THREE ESSAYS

IN

EMPIRICAL INDUSTRIAL ORGANIZATION

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ECONOMICS

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Matthew Shum

May 1998

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Frank A. Wolak

Prof. Frank Wolak
(Principal Advisor)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Timothy Bresnahan

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Peter C. Reiss

Prof. Peter Reiss

Approved for the University Committee on Graduate Studies:

Thomas Wasow

Abstract

This thesis consists of three empirical essays on topics in industrial organization. Chapter 1 introduces the three essays and describes their main results. In chapter 2, I measure the effects of advertising on households' choices among competing brands of differentiated products when households are habit persistent. In markets characterized by habit persistent consumers, an important role for advertising may be to encourage consumers to "switch" to newer, less familiar brands. This potentially *enhances* competition between competing brands by reducing household-level differentiation between brands that a household is and is not experienced with, contrary to the traditional Bainian arguments that advertising reinforces market power via product differentiation. Estimates from my household-level brand choice models for the breakfast cereals market imply that advertising not only promotes competition among existing brands by reducing the degree of product differentiation which arises due to experience, but also facilitates entry of new brands. Therefore Bain's argument that advertising is a "product differentiation barrier to entry" seems appropriate only in the presence of production or marketing economies of scope.

In chapter 3, which draws upon joint work with Andrea Coscelli, we empirically verify the existence of and quantify the extent of informational barriers to entry into the anti-ulcer drug market. Using an Italian dataset of complete prescription histories for 350 doctors in the Rome metropolitan area, we study the diffusion process of *omeprazole*, an anti-ulcer molecule which entered the market in 1990. Our estimates ascribe the gradual growth in prescriptions of *omeprazole* over the sample period is due largely to doctors' initial pessimism about *omeprazole*'s quality. Strikingly, we find that this uncertainty is resolved

by first-hand experience (actual prescriptions) rather than through marketing activities and other potential sources of information: this despite including a complete set of dummies for all the months in our sample in order to soak up aggregate effects.

In chapter 4, based on joint work with Han Hong, we derive an econometric model of a multi-round ascending (or English) auction. Much of the previous empirical work on auctions employing structural models have focused on *single-round* auctions. Furthermore, the model we develop accommodates *asymmetries* among the bidders in allowing their valuations for an object to be non-identically distributed. While the aim of this chapter is mostly methodological, we estimate our model using data from the spectrum auctions run by the U.S. Federal Communications Commission (FCC).

Acknowledgments

I feel particularly indebted to Frank Wolak, my principal advisor. He was a knowledgeable source of advice on all aspects — economical, statistical, and computational — of my research. As most graduate students realize, however, often the most important component of success is willpower; for that reason I am especially grateful to Frank for his constant encouragement and support. I could not have asked for a more patient or understanding advisor.

On repeated occasions, Peter Reiss prodded and pushed me towards coherence and clarity in making my economic arguments. He taught me how to write a paper. Tim Bresnahan taught me the *playfulness* of research: about gaining insights and developing economic intuition by allowing even antithetical ideas to combine and collide. Andrea Coscelli provided extensive comments on many drafts of chapter 2, and was a co-author of chapter 3. Furthermore, Andrea was a person who "got things done", and luckily some of it rubbed off on me. Han Hong's impressive knowledge of statistics and auction theory informed many aspects of chapter 4, of which he was a co-author. Finally, I have benefited from John Pencavel's advice since the beginning of my graduate school career, when he convinced me to pursue empirical research.

I gratefully acknowledge financial support from the Alfred P. Sloan Foundation.

This dissertation would simply not have been possible without my fellow graduate students and their valuable advice, feedback and encouragement. I especially thank my comrades-in-arms Klaus Desmet, Luis-Fernando Medina, and Fabiano Schivardi for their friendship.

Over twenty years ago, my grandfather taught me the Tang dynasty poem which begins (and I translate):

> Gentlemen, don't you see —
> The waters of the Yellow River — out from Heaven — how they rush out to the sea, never returning?

This thesis is dedicated to his memory.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> Therefore, my dear friend and companion, if you should think me somewhat sparing of my narrative on my first setting out, — bear with me, — and let me go on, and tell my story my own way: — or if I should seem now and then to trifle upon the road, — or should sometimes put on a fool's cap with a bell to it, for a moment or two as we pass along, - don't fly off, — but rather courteously give me credit for a little more wisdom than appears upon my outside; - and as we jogg [*sic*] on, either laugh with me, or at me, or in short, do any thing, — only keep your temper.[1]

This thesis consists of three empirical essays on topics in industrial organization. In Chapter 2, I measure the effects of advertising on households' choices among competing brands of differentiated products. In markets characterized by habit persistent consumers, an important role for advertising may be to encourage consumers to "switch" to newer, less familiar brands. This potentially *enhances* competition between competing brands by reducing household-level differentiation between brands that a household is and is not experienced with, contrary to the traditional Bainian arguments that advertising reinforces market power via product differentiation.

Since habit persistence is a household- or individual-level phenomenon, investigating these issues requires detailed microdata. Using a panel dataset collected from individual households' purchases of cereal in six large supermarkets in the Chicago metropolitan area, I estimate household-level models of brand choice among competing brands of cereal, in which a household's experience with a brand is measured by its recent purchases. In order to draw

[1] L. Sterne, *Tristram Shandy* [69], p. 10.

proper inferences regarding the effects of experience, I control for unobserved heterogeneity in household preferences which, as is well known (see Heckman [38]), can induce "spurious" brand loyalty in panel data models with lagged dependent variables.

Furthermore, I observe only households' purchases at any of the six supermarkets included in the study, not their complete purchase histories. Therefore I develop and implement a Probabilistic Nested Logit (PNL) demand model to control for the possibility that not all of a household's purchases (and therefore not all of its experiences with certain brands) are observed. The simulation methodology I develop to estimate this model is directly applicable to "probabilistic choice set" settings which accommodate the researcher's uncertainty about the size and composition of agents' choice sets (see Ben Akiva and Lerman [11] and Swait [72] for more details on these models).

Two main results emerge from this analysis. First, habit persistence is very prevalent: experience increases purchase probabilities and lowers price sensitivities. Secondly, advertising "mimics" experience by likewise increasing purchase probabilities and reducing price sensitivities — *but only for the inexperienced brands.* Therefore, advertising indeed encourages switching behavior by raising the appeal of the inexperienced brands relative to the experienced brands at the household-level. These findings imply that advertising not only promotes competition among existing brands by reducing the degree of product differentiation which arises due to experience, but also facilitates entry of new brands. Therefore Bain's argument that advertising is a "product differentiation barrier to entry" seems appropriate only in the presence of production or marketing economies of scope.

Chapter 1 focuses primarily on cross-sectional product differentiation: differentiation among a set of product at a given point in time. Past empirical research has also demonstrated the existence of product differentiation *over time* in the form of *first-mover advantage* in oligopolistic markets. Previous researchers (among them Schmalensee [65] and Bagwell [7]) have attributed first-mover advantage to agents' lack of information about the relative quality and attributes of new entrant products.

In chapter 2, which draws upon joint work with Andrea Coscelli, we empirically verify the existence of and quantify the extent of these "informational barriers to entry" into the anti-ulcer drug market, the largest therapeutic drug market worldwide. Using an Italian dataset of complete prescription histories for 350 doctors in the Rome metropolitan area which offers unprecedented detail into doctors' prescribing behavior, we study the diffusion process of a new anti-ulcer molecule. *Omeprazole*, patented by Astra, and marketed worldwide under the brand name LOSEC (but in the U.S. as PRILOSEC), became the world's best selling drug in 1996, grossing more than 2bn$, and gaining about half of the worldwide anti-ulcer market. However, this market ascendance was gradual; during the sample period, drugs based on *omeprazole* managed to garner only 20% of the Italian market, and even in mid-1995, it still clearly lagged behind the incumbent H_2-antagonist molecules. To what extent is this gradual growth in market share attributable to doctors' reluctance to prescribe the molecule, due to lack of information about its quality? Furthermore, how important is first-hand experience with *omeprazole* (i.e., actual prescriptions of the molecules to patients) in informing doctors about its quality, relative to other possible sources of information, such as marketing activities undertaken by the manufacturers of the new *omeprazole*-based drugs? To answer these questions, we specify a learning model in which doctors, initially uncertain about the quality differential between *omeprazole* and the incumbent molecules, update their beliefs about this differential after observing, through first-hand experience, noisy signals of it from patients to whom they have prescribed the molecule.

Estimates from the learning model attribute the gradual growth in prescriptions of *omeprazole* over the sample period to doctors' initial pessimism about *omeprazole*'s quality. While this is not unexpected, given the strong serial correlation in prescriptions of *omeprazole* at the doctor-level observed in our data, we also find that doctors' uncertainty about *omeprazole*'s quality is overwhelmingly resolved by first-hand experience (actual prescriptions) rather than through marketing activities and other potential sources of information (such as journal publications and word-of-mouth among doctors) at the aggregate level. Remarkably, these results obtained even though we "stacked the cards" against them by including a full set of time-period dummies to soak up any factors left unexplained by either the learning process or the included covariates. In other words, we find that the gradual rise in market share in our data is driven by *doctor-level*, rather than aggregate (across all doctors) serial correlation in prescriptions of *omeprazole* . This is strong evidence in favor of the learning model in which, similarly, movements in aggregate market share are driven by doctor-level learning rather than by aggregate effects.

While chapter 2 focuses on the decision problem which confronts a doctor about whether or not to prescribe a new drug to a given patient given uncertainty about the efficacy of this new drug, chapter 3 focuses on a similar problem — valuing an object with uncertain quality — not simply from the perspective of a single agent, but rather from the perspective of a group of agents who are competing with each other to gain the object of unknown quality at the lowest cost. In chapter 3, based on joint work with Han Hong, we derive an econometric model of a multi-round ascending (or English) auction. Much of the previous empirical work on auctions employing structural models (two examples are Paarsch [61] and Laffont, Ossard and Vuong [46] have focused on *single-round* auctions. Furthermore, the model we develop accommodates *asymmetries* among the bidders in allowing their valuations for an object to be non-identically distributed.

While the aim of this chapter is mostly methodological, we estimate our model using data from the spectrum auctions run by the U.S. Federal Communications Commission (FCC). Our model captures the multi-round aspect of the FCC auctions, but we have abstracted away from the synergies that may result from the simultaneous auctioning off of multiple objects; furthermore, the flexible eligibility rules of these auctions are at odds with the "irrevocable dropout" assumption of the Milgrom-Weber [57] model upon which our auction model is based. Therefore, we present the empirical work more to illustrate and suggest solutions to problems which arise in estimating this model in practice than to empirically test economic hypotheses concerning the FCC auctions.

Despite their apparent disparity in focus, my emphasis throughout is on measuring the dynamics of demand in markets with either a limited number of sellers or buyers. Chapters 2 and 3 focus on analyzing the nature of and determinants of demand in two oligopolistic markets: the breakfast cereals and Italian anti-ulcer drug market, respectively. The profitability of oligopolistic markets are often sustained by means of product differentiation (offering a product which is only imperfectly substitutable with the products offered by competing sellers). Therefore sizeable effort will be devoted to investigating and quantifying dimensions of differentiation in these two markets: the role of advertising and brand loyalty in the case of breakfast cereals, and uncertainty about the quality of a new drug vis-a-vis existing drugs in the Italian anti-ulcer drug market. Chapter 4, on the other hand, focuses on buyer behavior in a "market" where the buyers are few: the bidders vying for an object in an ascending auction. Here, the market mechanism can serve to "pool" private information that individual buyers may have about the value of the product. This possibility of information acquisition during the bidding process is one main feature of the auctions modeled in this chapter.

Chapter 2

Advertising and Competition with Habit Persistent Consumers: Estimates from a Probabilistic Nested Logit Model

Does advertising inhibit or promote competition in differentiated product markets? Economists have long debated this question. The contention that advertising fosters product differentiation among otherwise very similar brands forms the basis of Bain's view that advertising inhibits competition by reinforcing a "product differentiation barrier to entry" ([8][1]). Some cross-industry empirical work (summarized in Ferguson[32]) tends to support this view by reporting a positive relationship between advertising intensity, concentration and profitability across industries. More recently, Sutton [71] demonstrates that a robust negative relationship between advertising and industrial concentration arises in markets where advertising functions as an "endogenous sunk cost" (i.e., the optimal amount of advertising expenditure varies with the number of firms in the market) — thus grounding the Bainian

[1] pg. 114: "product differentiation is propagated by [...] advertising and sales-promotional efforts designed to win the allegiance and custom of the potential buyer."

hypothesis in a rigorous game-theoretic framework.

All the research cited above suggests a *negative* relationship between advertising and competition. Very different implications arise in the "informative advertising" literature (beginning with papers by Stigler [70] and Nelson [59]; also summarized in Ferguson [32], ch. 2). In these models, *ex ante*[2] differentiation (i.e., low substitutability) exists among the competing products due to consumers' lack of information about the existence and attributes of the competing brands. The role of advertising is to provide this information thus *reducing* informational product differentiation and promoting competition.

On the surface, the informative advertising argument appears less relevant in mature product markets (such as the one for breakfast cereals, which is the focus of this chapter), where very little national advertising can seriously be classified as "informative" since the most highly-advertised brands have been around for decades. This does not mean that the informative advertising literature is irrelevant in these markets: generally, the *mechanism* whereby advertising promotes competition in the informative advertising models — reducing *ex ante* differentiation among the competing brands — can also arise in situations in which advertising's role is not explicitly informative.

This chapter focuses on the competitive effects of advertising in markets where consumer preferences are characterized by **habit persistence** (or "brand loyalty", in marketing jargon) a heightened willingness to pay for brands that a consumer has recently experienced.[3] With habit persistent consumers, experience leads to differentiation (i.e., lower substitutability) between experienced and inexperienced brands in the same fashion as with imperfectly informed consumers: if Mrs. Schnurbart is not perfectly informed about

[2]before consumers are exposed to advertising

[3]Habit persistence, as defined here, is a household-level phenomenon which depends only on whether a household has had previous experience of the brand; its effects do not differ across brands. For example, provided that it has tried both Cheerios and Frosted Flakes recently, I define household h as being "habituated" to both. This is not the same as *brand equity*, which describes the attractiveness of *particular brands*, regardless of households' previous experience. The latter is a measure of how *vertical differentiated* a particular brand is: a brand with high "equity" is thought to have high quality, roughly speaking. In my model (as explained below), equity/quality differences across brands are captured by brand dummies, which are assumed to be the same over all households.

Furthermore, note that experience does not "cause" habit persistence; I assume that habit persistence is an intrinsic aspect of consumer preferences. Reliance on previous experience as a guide to future purchases is simply the *manifestation* of this intrinsic habit persistence.

the range of products available, there is low substitutability between brands which she is and is not informed about. Analogously, if Mrs. Schnurbart is very habit persistent — if she becomes less price sensitive to brands that she has recently tried — then there is a low degree of substitutability between brands which she has and has not recently tried.[4] In both cases, a potentially important role of advertising is to encourage "switching" behavior at the individual household level, to persuade households to try (by increasing their willingness-to-pay for) brands that they have not recently experienced. The ensuing advertising-engineered increase in substitutability among the available brands at the household level can potentially translate into *positive* effects on competition at the market level.

Shapiro [66] has succinctly summed up these generally pro-competitive effects of advertising under the rubric that advertising "substitutes for experience"; he also uses the "brand loyalty" moniker (emphasis mine):

> advertising serves as an entry barrier when it is a complement to experience in creating brand loyalty, while it promotes entry when it *substitutes for experience* in the production of brand loyalty.

Advertising substitutes for experience by mimicking its effects. In the informative advertising case, advertising provides consumers with information they would otherwise have to gain through experience by costly sampling. Similarly, habit persistent consumers should be drawn to highly-advertised brands in the same fashion as if they had recently experienced those brands: this will be reflected in a propensity to "switch" from experienced brands to inexperienced, but highly-advertised brands.

The goals of this chapter are therefore twofold: (1) to quantify the extent of habit persistence — the effects of experience, namely — in a differentiated product market; and (2) to test whether advertising substitutes for (or mimics the effects of) experience by

[4]Chapter 3 deals explicitly with measuring quantifying information product differentiation in the Italian anti-ulcer drug market.

promoting switching behavior.[5] The results shed light on advertising's effects on *inter-brand* competition in this market, when households' preferences are characterized by habit persistence.

The empirical analysis in this paper focuses on the breakfast cereals market. The huge amounts of advertising spent in this market cannot be labeled "informative" since older, established brands remain highly-advertised, despite the large numbers of new brand introductions. Why do *Cheerios* and *Frosted Flakes* continue to be advertised so intensively? This chapter focuses on one possible answer to this question: that if consumers' preferences are characterized by habit persistence, advertising — even for as well-known a brand as *Cheerios* — must be maintained in order to convince consumers habituated to rival brands to switch.

Since habit persistence is a household- or individual-level phenomenon, answering the questions raised earlier requires detailed microdata. I employ a panel dataset collected from individual households' purchases of cereal to estimate household-level models of brand choice among competing brands of cereal, in which a household's experience with a brand is measured by its recent purchases. In order to test whether advertising substitutes for, or mimics, experience, I focus on the tradeoff between advertising and experience, investigating how advertising's effects differ across households depending on their experience with a brand.

In order to draw proper inferences regarding the effects of experience as measured by past purchases, I control for unobserved heterogeneity in household preferences which, as is well known (see Heckman [38]), can induce "spurious" habit persistence in panel data models with lagged dependent variables. Furthermore, my dataset includes households' cereal purchases at only six Chicago-area supermarkets, not their complete purchase histories. Therefore I devise an **probabilistic nested logit (PNL)** brand choice model to control

[5] I do not attempt to explain *why* advertising substitutes for experience; as I noted before, the information interpretation seems inappropriate in the breakfast cereals market, where most of the brands have been around for years. I just test whether advertising has the same types of effects (same sign or magnitude) as experience.

for the possibility that not all of a household's purchases (and therefore not all of its experiences with certain brands) are observed. The simulation methodology I develop to estimate this model is directly applicable to "probabilistic choice set" settings which accommodate the researcher's uncertainty about the size and composition of agents' choice sets (see Ben Akiva and Lerman [11] and Swait [72] for more details on these models).

Preview: main findings Two main results emerge from my analysis. First, habit persistence is very prevalent: experienced households have higher purchase probabilities and lower price sensitivities. For one of the preferred specifications, being experienced with a brand lowers a household's own-price elasticity for an average brand from -5.6 to -2.2,[6] while increasing the purchase probability from 0.2% to 9.5%.

Second, I find that, similar to experience, advertising reduces price sensitivities — *but only for the inexperienced brands.* For the experienced brands, advertising actually *increases* price sensitivities. Both these effects imply that advertising indeed encourages switching behavior, by raising the appeal of the inexperienced brands relative to the experienced brands.

One competitive implication of these results is that advertising is potentially much more useful for new brands, in the sense that *ceteris paribus*, a marginal increase in advertising expenditures would increase sales of a new brand more than sales of an existing brand. Therefore, in addition to increasing competition among existing products by encouraging switching behavior, advertising may also play an important role by facilitating the entry of new brands. Thus my empirical results raise some doubt about Bain's characterization of advertising as a "product differentiation barrier to entry". Rather, they provide empirical confirmation for Shapiro's [66] above-cited theoretical argument that *experience* is the actual barrier to entry in markets characterized by habit persistence, and advertising which "substitutes for experience in the creation of brand loyalty"[7] actually lowers this barrier to entry.

[6] For the sake of brevity, I will always refer to reductions in the *magnitude* of the own-price elasticity as reductions in the elasticity itself.

[7] which creates "synthetic" brand loyalty towards brands which a household has not actually experienced

My analysis focuses on households' choice among different *brands* in this market, abstracting away from their choices of *quantities*.[8] This limits my ability to draw welfare implications from my results because the welfare benefits accruing to any specific household from (say) a given change in advertising depends not only on the brand that it purchases, but also on the amount purchased.[9] Nevertheless, my results have shed light on the effects of advertising on *inter-brand* competition in a market where individual preferences are characterized by habit persistence — as much of the empirical marketing literature shows, habit persistence appears to be an important characteristic of demand in many consumer nondurables markets.

2.1 The breakfast cereals market: background and data

2.1.1 Advertising in the cereals market

The breakfast cereals market is a particularly appropriate one to examine the issues raised previously because it seems a likely one in which these effects could be present.

First of all advertising in this market seems extraordinarily high, even by the standards of the already high-advertising food manufacturing sector: advertising intensity (as measured by the ad-to-sales ratio) for the Grain Mills Products industry (SIC 2040, the bulk of which is cereals) is about 1.2 times the average value for the food sector, and about 3.5 times higher than the average value for all industrial sectors (see figure 2.1).

At the same time, there are an incredible number of brands of cereals available at any one time (218 distinct brands appear in my dataset, with an average of around 110 available at a supermarket during any given week). These two characteristics of the industry — high advertising intensity, and substantial product differentiation — by themselves tend to justify

[8]This is because it seems reasonable to assume that, as a competitive weapon, advertising for cereals serves to influence households' brand choice more than their quantity decisions.

[9]Many of the discrete choice models defined at the individual level but aggregated to the market level and subsequently estimated with market-level data (e.g., Stern [68] and Nevo [60]) do not explicitly model an individual's *joint* choice of brand and quantity either; they implicitly assume that, for example, household A buying 2 boxes of cereal X is equivalent to two households with the same characteristics as household A, each buying 1 box of cereal X.

Figure 2.1: Advertising indicators for selected industries

Industry	**Ad. expns**[a] ($ mills)	**Ad-sales ratio**[b]	**Ad-margin ratio**[c]
Food and Kindred Products (SIC 2000)	4,033.1	7.7	19.9
of which:			
Grain Mill Products (SIC 2040) (*includes:* RTE Cereals)		9.2	19.6
Bakery Products (SIC 2050) (*includes:* Cookies/Crackers)		9.0	41.0
Beverages (SIC 2080) (*includes:* Carbonated Soft drinks)		5.8	10.8
Average: all industries		2.7	
Total: all industries	59,887		

[a]Source: Leading National Advertisers. Data for 1995.
[b]Source: Schonfeld and Associates. Data for 1996. Advertising expense divided by net sales (after discounts and allowances) as a percentage. Indicator of advertising intensity.
[c]Source: Schonfeld and Associates. Data for 1996. Advertising expense divided by gross margin (net sales-cost of goods sold) as a percentage.

the traditional Bainian arguments that advertising sustains product differentiation among the competing brands.

However, this market is also characterized by extensive new brand entry: Hausman [37] notes that approximately 190 brands were introduced in the 1980-92 period (pg. 213). The high advertising expenditures could therefore be justified for informational reasons alone, since in this industry the introduction of a new brand is usually surrounded by a flurry of promotional activity.[10]

[10]Hausman also cites industry estimates that $20-40 million must be spent in the initial year to advertise and promote a new brand.

Table 2.1: Brand characteristics: Summary statistics

	Name	Avg transaction Price	Avg Shelf Price	Avg Ad Expn	Nat'l Market Share	In-sample Market Share	Exist in 1983? In 1988?	Sgmnt
1	KG[11] Corn Flakes	1.81[12]	1.95[13]	7.109[14]	5.1[15]	5.67[16]	xx[17]	Fam
2	GM Cheerios	3.16	3.47	7.287	4.8	4.38	xx	Fam
3	KG Rice Krispies	2.96	3.20	6.034	3.8	4.04	xx	Fam
4	KG Frosted Flakes	2.52	2.68	7.867	4.5	3.82	xx	Fam
5	KG Raisin Bran	2.34	2.50	5.591	3.2	2.73	xx	Fam
6	GM Total	3.61	4.04	3.926	1.8	2.36	xx	Adult
7	GM HoneyNut Cheerios	3.14	3.41	4.030	2.7	2.26	xx	Fam
8	KG Special K	3.48	3.78	3.531	1.3	2.16	xx	Adult
9	PT Grape Nuts	2.14	2.29	6.740	2.9	2.12	xx	Adult
10	NB SpoonSize ShdWt	2.81	3.05	0.025	1.2	2.08	xx	Adult
11	QK 100% Natural	2.24	2.55	1.612	1.0	1.96	ox	Adult
12	KG Frosted Mini Wheats	2.62	2.75	6.106	2.8	1.84	xx	Adult
13	KG NutriGrain	2.87	3.10	2.508	0.8	1.55	xx	Adult
14	KG Mueslix	3.31	3.58	1.975	0.8	1.53	oo	Adult
15	GM Wheaties	2.55	2.86	2.257	1.4	1.52	xx	Fam
16	PT Raisin Bran	2.23	2.57	4.361	1.9	1.46	xx	Fam
17	RL Muesli	3.34	3.93	0.215	0.4	1.26	oo	Adult
18	KG Corn Pops	3.51	3.69	3.198	1.0	1.46	xx	Fam
19	GM Raisin Nut Bran	2.98	3.22	1.659	1.1	1.35	ox	Adult
20	GM Basic 4	3.27	3.63	2.510	0.8	1.31	oo	Adult
21	GM Cocoa Puffs	3.46	3.67	2.097	0.6	1.28	xx	Kids
22	GM Golden Grahams	3.24	3.54	2.953	1.0	1.24	xx	Kids
23	GM Cinn. Toast Crunch	3.36	3.56	2.963	1.2	1.23	ox	Kids
24	KG Froot Loops	3.53	3.76	3.110	1.9	1.20	xx	Kids
25	KG Low Fat Granola	2.68	3.10	2.327	0.9	1.17	oo	Adult
26	GM Trix	3.96	4.22	3.236	1.2	1.13	xx	Kids
27	GM Triples	2.33	2.80	3.036	0.9	1.12	oo	Adult
28	KG Crispix	3.28	3.49	3.225	1.2	1.12	xx	Adult
29	GM Kix	3.67	3.93	3.801	1.2	1.08	xx	Kids
30	GM Lucky Charms	3.45	3.72	3.079	1.4	1.08	xx	Kids
31	GM AppleCinn. Cheerios	3.02	3.35	3.120	NL	1.06	oo	Fam
32	KG Cracklin Oat Bran	3.19	3.51	2.279	0.9	1.06	xx	Adult
33	NB Big Biscuit ShdWt	2.79	3.05	0.000	0.8	0.99	oo	Adult
34	PT Honey Bunches of Oats	2.85	3.18	3.749	1.1	0.95	oo	Adult
35	PT Great Graines	2.90	3.43	2.648	0.5	0.89	oo	Adult
36	GM Otml Raisin Crisp	2.71	3.04	1.641	0.5	0.97	ox	Adult
37	QK Oat Squares	2.43	2.71	1.472	0.8	0.94	ox	Adult
38	RL Rice Chex	3.40	3.53	0.875	0.8	0.89	xx	Adult

[11] KG: Kelloggs GM: General Mills PT: Post (Phillip Morris) RL: Ralston QK: Quaker Oats
[12] prices in $/lb.
[13] prices in $/lb.
[14] quarterly expns, $million. **Source:** Leading National Advertisers [49]. Avg'd over 1991:ii—1993:ii
[15] Shares of national cereal volume, 1992. **Source:** IRI [41]. NL: brand not listed in source.
[16] Share of total in-sample purchases. **Source:** Author's calculation from Stanford Basket Dataset.
[17] xx: exist in both years; ox: exist in 1988, not in 1983; oo: exist in neither year. **Source:** IRI [41].

Table 2.1: Brand characteristics: Summary statistics (*continued*)

	Name	Avg transaction Price	Avg Shelf Price	Avg Ad Expn	Nat'l Market Share	In-sample Market Share	Exist in 1983? In 1988?	Sgmnt
39	GM Total Raisin Bran	3.00	3.50	1.874	0.4	0.89	ox	Adult
40	KG Product 19	3.38	3.70	1.408	0.4	0.89	xx	Adult
41	KG Apple Jacks	3.64	3.91	1.465	0.7	0.84	xx	Kids
42	QK Capt Crunch	2.55	2.86	1.714	1.8	0.83	xx	Kids
43	NB Shredded Wheat	2.82	3.00	2.925	0.5	0.80	xx	Adult
44	PT Fruity Pebbles	3.21	3.48	1.710	0.8	0.83	xx	Kids
45	GM Clusters	3.14	3.52	1.425	0.9	0.78	ox	Fam
46	KG Cinnamon MiniBuns	2.75	3.14	0.002	0.8	0.76	oo	Fam
47	KG Double Dip Crunch	3.01	3.52	1.454	0.6	0.73	oo	Adult
48	GM MultiGrain Cheerios	3.34	3.74	2.520	NL	0.75	oo	Fam
49	PT Honeycomb	3.40	3.67	2.567	0.7	0.74	xx	Kids
50	QK Popeye	1.77	1.77	0.000	NL	0.67	oo	Kids
51	basket of all other brands	2.68		0.645[18]	34.3	24.29		

But this appears not to be the case. Table 2.1 contains average quarterly advertising expenditures for the top fifty brands in the cereals market in the sample 1991-93 period (column 5), as well as information on whether each brand had already existed in 1983 and 1988 (column 8). Clearly, the most highly-advertised brands are well-established brands, such as *Cheerios* and *Frosted Flakes*, which raise some doubt as to the suitability of the information explanation for this industry's high advertising intensity.

These considerations motivate the main question addressed in this chapter: why do established brands continue to advertise so intensively? The traditional Bainian answer is that advertising maintains the differentiation among the products which allows them all to remain profitable in this market. In marketing terms, advertising and other promotional activity transforms a product from a mere "commodity" into a distinct "brand".[19]

This paper tests between this explanation and the alternative described earlier: that if consumers' preferences are characterized by habit persistence, advertising — even for as well-known a brand as *Cheerios* — must be maintained in order to convince consumers habituated to rival brands to switch. The competitive implications for these two explanations

[18]sum of average quarterly advertising expenditure for *all* the non top fifty brands

[19]This view is apparent in a securities analyst's comment on Kellogg's decision last year to reduce ad spending in order to pay for its price cuts: "If the industry goes away from building quality and image, that's a mistake and could lead to commodization of the category" [21].

differ: the Bainian explanation implies that advertising reduces competition by increasing product differentiation, while with habit persistent consumers advertising reduce differentiation *at the individual consumer level*, which could potentially promote competition at the market level.

2.1.2 Data

Since habit persistence is an individual-level phenomenon, aggregate market share data similarly to that used by Hausman [37] and Nevo [60] in their previous studies of the breakfast cereals industry is inappropriate for my purposes. In my empirical work, I employ detailed household-level purchase history data. The Stanford Basket Dataset (compiled from the database of Information Resources, Inc., a leading market research firm) is a household panel which tracks the purchases of 1010 households in six supermarkets in the Chicago metropolitan area on a weekly basis over a two-year period (1991-1993). The purchase data are recorded by scanners at supermarket checkout counters. Demographic characteristics for each household are also included in this dataset, allowing me to explicitly account for across-household differences in preferences, which must be controlled for when modeling brand choice probabilities in horizontally-differentiated markets like the one under study.

This purchase data is matched with quarterly aggregate (i.e., national) brand-level advertising expenditures data from Leading National Advertisers.[20] Given the brand-level nature of my data, therefore, I focus on measuring advertising's effects on *brand choice* in the cereals market.[21]

[20]The ideal advertising variable to use in parameterizing household h's probability of purchasing brand i on purchase occasion t would have been household h's *exposure* (measured as number of messages, such as TV commercials, magazine ads, etc.) to brand i's advertising on or immediately preceding purchase occasion t. However, I do not have covariates in my data which directly control for this type of variation across households.

[21]I abstract away from advertising's possible effects on a household's decisions whether or not to purchase cereal, and how much, since the first order determinants of these decisions are most likely household demographics (such as family size) rather than advertising.

Furthermore, using advertising dollars instead of number of advertising messages assumes that the "price" of a message is the same over all brands, and over all time. Aggregation to a national level, as well as the fact that most cereals primarily employ broadcast advertising, justify this assumption. Most variation in broadcast advertising costs is over the different times of the day (the "dayparts").

There are hundreds of brands of breakfast cereal (218 in my sample), and the largest among them do not enjoy much more than a 5 percent market share (see column 7 of table 2.1). Given this diffuseness, I model households' choices among 51 brands in this market, with brands 1–50 being the fifty brands with the highest shares of total purchases in the basket dataset, and the 51st brand being a composite of all the other (non top-fifty) brands. Since stores vary in the non top-50 brands that they carry, the composition of this 51st brand therefore varies across households as well as over time, depending on which stores the households choose to shop in.

Table 2.1 presents summary characteristics for the 51 brands of cereals used in my analysis. The top fifty brands accounted for just over 75% of all purchases in my data. In my sample period, the average store carried about 64 non top-fifty brands during a given week; this indicates that, the average supermarket cereal aisle contains over a hundred brands of cereal. I classified the top fifty brands into three segments, taking as a guide the classification schemes in Hausman [37]: FAMILY (13 brands), ADULT (25 brands), and KIDS (12 brands).[22] Furthermore, comparing columns 6 and 7 of table 2.1 shows that in-sample market shares (calculated from total purchases) are by and large similar to national market shares (calculated over total volume).

Prices for each brand on any purchase occasion are not straightforward to construct given my data since the possibility of coupon discounts means that oftentimes households' *transactions prices* (price actually paid at the checkout counter) are substantially lower than the *shelf prices*. In my data, I observe shelf prices for all brands which were sold in a store for any given week[23], but transactions prices only for purchased alternatives. The missing datum is what a household *would have* paid for the unpurchased alternatives.

[22]Note that that family cereals are advertised the most but are cheaper than both adult and kids cereals. While the advertising numbers include expenditures on ten media, almost all of the advertising dollars were spent on broadcast (television) advertising, about evenly divided up between national and spot (local) television. Interestingly, most family cereals are advertised on national television, while adult and children cereals are advertised on spot TV. This may reflect "targeting" on the part of the cereal manufacturers, to the extent that spot TV allows finer tuning towards target demographics (on a geographic basis, at least) than national TV. However, this intuition was not confirmed by my telephone conversations with various industry observers. See Media Dynamics, Inc, [55] for more information on television advertising.

[23]Furthermore, I assumed that brands which were not sold during any given week in a given store were in fact not available. This rarely happened among the top fifty brands.

There are two possible ways of calculating prices for estimation. First of all, I can use store- and week-specific **shelf prices**, ignoring possible price discounts. Secondly, I can average over observed transactions prices in order to derive store- and week-specific **average transactions prices** for each brand. These prices are meant to capture the prices that a household *would have* paid for a given brand. The third and fourth columns in table 2.1 indicate that average transactions prices are markedly lower than shelf prices, often on an order exceeding 10%. Both sets of prices would tend to understate the price sensitivities of households, since both would understate households' response to price discounts.[24] Below I present mostly results from models estimated with average transactions prices, since the results using shelf prices do not differ much qualitatively or quantitatively.

Purchase data: summary statistics Households frequently made more than one shopping trip within a given week. I aggregated all shopping activity up to a weekly basis in order to abstract from frequency of shopping issues, which I expect to be determined largely by household demographics. In other words, *I model households' weekly brand choices conditional on their making a shopping trip.* The sample selection involved in ignoring weeks where households do no shopping introduces bias into the parameter estimates of my brand choice model only if there is systematic correlation between the frequency of purchase and the brands which are bought — in other words, if "frequent shoppers" always tend to buy certain brands. In abstracting away from the frequency of purchase issue, I am implicitly assuming that *conditional on the included covariates*, no such correlation is present.

Furthermore, there are weeks in the sample in which a household makes multiple purchases, i.e., purchases of more than a single brand. I abstract away from the multiple purchases dimension of brand choice, and model each of these within-week purchases as *i.i.d.* observations. This was done mainly for computational convenience.[25]

[24] Using shelf prices assumes that households are not aware of any price discounts; using transactions prices assumes that households are aware of all price discounts. Both may underestimate price sensitivity in a world where, on each purchase occasion, households may be only aware of price discounts for the brand which they end up buying.

[25] Households purchased more than one brand in 27.1% (100-72.9) of all (household-weeks) in which purchase occurred. The following table characterizes the empirical distribution of the number of distinct brands purchased per week, conditional on purchase. Each observation is a (household-week) in which

Following this scheme, there are 88,419 observations in my dataset. Table 2.2 contains summary statistics of the data used in the estimation. The dependent variable is a discrete variable which equals 0 if no purchase occurred, and equals i if brand i was bought. As the summary statistics on the POSPURC variable shows, nonpurchase occurs for about 65% of the observations.

2.1.3 Representativeness of sample households

1. Household demographics The households in the basket data sample tend to be more elderly then an "average" American household, as indicated in figure 2.3. Apparently IRI chose these households in order to minimize attrition during the two-year sampling period. For both males and females, the mode of the distribution of age ranges is very high: between 65 and 99 years. (In the basket data, only the age range of the male and/or female household heads are given, not the actual age.) Furthermore, 30.9% of the households do not contain a male: presumably, these are widow households. Comparable figures for the 1992 Consumer Expenditure Survey (hereafter CEX) diary module are given in the last column: the mode of the age distribution of the household head in the CEX households is much lower, in the [35,44] range.

However, just over 70% of the households in my sample contain children under the age of 18; this is captured by the dummy variable YGCHIL. In this instance, comparison with the CEX shows that there are no strong biases in my sample with respect to households with and without children, despite the larger age of the household heads.

purchase was observed. NUMPURC is the number of distinct brands bought in a given week.

NUMPURC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	12155	72.9	12155	72.9
2	3356	20.1	15511	93.0
3	863	5.2	16374	98.2
4	211	1.3	16585	99.4
5	65	0.4	16650	99.8
6+	28	0.1	16678	100.0

See also footnote 32 below.

Figure 2.2: Household data: Summary statistics

Variable	Definition	N	mean	StdDev	min	max
Household demographics						
YGCHIL	=1 if kids under age 18 present[a]	1010	0.29	0.45	0	1
FAMSIZE	Number of persons in household[b]	1010	2.61	1.37	1	6
LOGINC	Log(family income/10000)	1010	0.97	0.81	-0.69	2.08
Household-level purchase summaries						
WEEKOBS	Weeks in which shopping trip occurred	1010	78.26	14.96	34	104
WEEKPURC	Weeks in which purchase of cereal purchase occurred	1010	20.56	15.92	0	99
WEEKRAT	= WEEKPURC / WEEKOBS	1010	0.270	0.206	0	0.980
Time-varying household covariates						
POSPURC	=1 when a purchase was observed	88419	0.341	0.474	0	1
PREVPURC	=1 when cereal purchase occurred in previous week	87493	0.21	0.41	0	1
Miscellaneous						
N_{51}	# brands included in composite 51st brand	823[c]	63.64	12.94	38	100

[a]The data do not allow me to distinguish between households with only very young children and those with only teenagers

[b]Topcoded at 6

[c]Number of (store-weeks) observed in data

Figure 2.3: Basket dataset oversamples older households

Age Range	Basket Data		1992 CEX Diary Data
	% hholds with female in age range	% hholds with male in age range	% hholds with head in age range
hline not given	0.006	0.009	
[18,29]	0.015	0.010	0.174
[30,34]	0.057	0.026	0.112
[35,44]	0.197	0.142	0.221
[45,54]	0.186	0.132	0.161
[55,64]	0.173	0.149	0.114
[65,99]	0.314	0.224	0.218
no male		0.309	
no female	0.053		

2. Cereal purchase frequencies The middle panel in figure 2.2 gives some idea of the frequency of cereal purchases for the households in the sample. The WEEKOBS variable shows that the average household made shopping trips for 78.3 weeks (out of a maximum 104). On average, 20.6 of these weeks contained purchases of one or more brands of cereal (as indicated by the WEEKPURC variable). The WEEKRAT variable indicates that households purchase cereals in just 27% of the weeks in which they went shopping. This implies, on average, only one cereal purchase per month, which seems a very low figure.

Given the sampling scheme of my data, there are two possible explanations for this low cereal purchase probability: (1) older households purchase less cereal; and (2) the purchase data is incomplete because it includes the sample households' purchases at only six supermarkets. Comparing the sample purchase probabilities for the basket data with those from the 1992 CEX (diary module) suggest that the second explanation is more likely. Figure 2.4 presents sample weekly cereal purchase probabilities for both datasets. The overall mean of 0.198 in the basket data is not much more than half of the CEX mean of 0.376. This same underestimation trend follows even when I break down the households into whether they have children, or by family size. Furthermore, since many of the basket households do not have a male in them, I also calculate these numbers for female-headed households in the CEX. The same trend persists. Finally, I look at weekly purchase

Figure 2.4: Weekly cereal purchase probabilities from Basket Data and 1992 Consumer Expenditure Survey (Diary module)

FAMSIZE	Weekly PurcProb	StError
Basket data: over all households		
All	0.198	0.0012
2	0.199	0.0021
4	0.237	0.0033
1992 CEX[a]		
Over all households		
All	0.377	0.0046
2	0.360	0.0082
4	0.528	0.0121
Elderly[b] households		
All	0.531	0.0137
2	0.404	0.0191
4	0.521	0.0533
Elderly fem-headed households		
All	0.340	0.0219
2	0.366	0.0392
4	0.684	0.0969

[a]The CEX diary survey includes information on *all* cereal purchases over a two-week period. For each household, I created two indicator variables describing whether or not this household purchased cereal at all during a given week. Purchase probabilities were calculated by summing over these indicator variables over all households, and dividing by the total number of (household-weeks) in the CEX diary data for 1992.

[b]Age of household head ∈ [55,64]

probabilities in the CEX by age cells, since we have seen before that the basket sample is biased towards older households. Nonetheless, in all cases, the estimated weekly purchase probabilities are still higher in the CEX than for basket data households. Just consider the cereal purchase frequencies for 2-member households: in my sample, the weekly purchase frequency for these households is 19.9%; this is much lower than the CEX estimate for all households (36.0%), older households (40.4% for household heads in the 55-64 age range), female-headed households (33.9%), as well as older female-headed households (36.6%, again for households heads in the 55-64 age range). These comparisons suggests that omitted purchases are causing the low cereal purchase frequencies observed in the basket data.[26]

Unfortunately my data were not supplemented with a household spending diary, so there is no way to gauge the extent of these omissions. Unobserved purchases would lead to measurement error in the experience variable, which is defined in terms of a household's recent purchases of a brand. Measurement error introduces biases into the estimates of the model coefficients only when correlation exists between the measurement error (which is subsumed into the model's disturbance term) and the included covariates. This is the case here, since clearly the determinants of unobserved as well as observed purchases are one and the same (namely, brand and household characteristics). This possible bias in the coefficient estimates motivates my development of the probabilistic nested logit (PNL) model below, which controls for this measurement error by explicitly parameterizing the probability of unobserved purchases as a function of household and brand characteristics.

Experience and habit persistence in brand choices I measure experience using an indicator variable PASTUSE_{iht} which is equal to one if household h bought brand i in the 12 weeks preceding purchase occasion t. Implicit in the use of this variable is the assumption that experience has no effect on brand choice after a 12-week "recall period" — due either to households' forgetting their experiences with a brand, or their preferences' changing in

[26]Similar comparisons by age and gender cells (not reported here) also show that households in the basket dataset spend less on nondurables expenditures on a weekly basis than households in the CEX diary module. This again suggests that the basket data is missing some shopping trips.

such a way that past experience is not valuable for current evaluations of a brand's utility.[27] A recall period of 12 weeks was chosen as a compromise between the length of the sample period and the length of the horizon over which households could be reasonably assumed to "forget" their previous experiences. Furthermore, exploratory probit regressions (results not reported here) which used PASTUSE to predict brand purchase probabilities indicated little qualitative differences in the results when varying the length (in weeks) of the recall period.

In lieu of summary statistics on the PASTUSE variables, I have done calculations (in figure 2.5) which show (1) how often households purchased inexperienced (i.e., PASTUSE=0) brands, and (2) on average, the number of brands which were experienced (i.e., PASTUSE=1) only for purchase occasions when a top fifty brand was purchased.[29] The w

Figure 2.5: How often do households buy inexperienced brands?

w	%INEXP	Avg. #Brands tried in last w weeks	NumObs
2	82.7	0.79	22363
4	70.8	1.49	21946
8	57.6	2.52	21136
12	50.5	3.32	20200
24	38.5	5.07	17560

Computed using all purchases of a top-fifty brand in my sample

column lists different lengths for the "recall period". The "%INEXP" column gives the percentage of observed purchases which represent purchases of brands which were inexperienced on the purchase occasion. Inexperienced brands are defined as those which were *not* consumed in the last w weeks. The row in which $w = 12$ corresponds to the 12-week recall period used in the definition of the PASTUSE indicator variable which I employ for the

[27] A less extreme approach would be to allow experience to fade over time, as in the "brand loyalty" models in marketing[28] in which experience is modeled as a stock variable which decays over time. However, initial conditions pose nontrivial problems in these models, as does determining the appropriate "decay rate" for the "stock" of experience.

Furthermore, the PNL model developed below to control for unobserved purchases relies crucially on the definition of experience as a dichotomous (0-1) indicator.

[29] I limit myself to observations where purchase of a top fifty brand occurred because the PASTUSE variable is defined slightly differently for the composite 51st brand: a household is defined as being "experienced" with this composite brand if it has consumed *any* non top fifty brands in the last w weeks.

analysis in this chapter. For this definition, 50.5% of purchases were of brands which were not purchased in the previous 12 weeks, and households purchased on average 3.32 brands in the 12 weeks preceding any given purchase occasion.

Furthermore, despite the large number of available brands, most households purchase few of these. In the two-year sample period, the average household in my sample purchases only 8.20 out of the top fifty brands. The mode of the empirical distribution of the number of different brands purchased by a household during the sample period (figure 2.6) occurs much lower, at only 3 brands. Indeed, households in general do not try many brands, and they tend to purchase the same brands over and over.

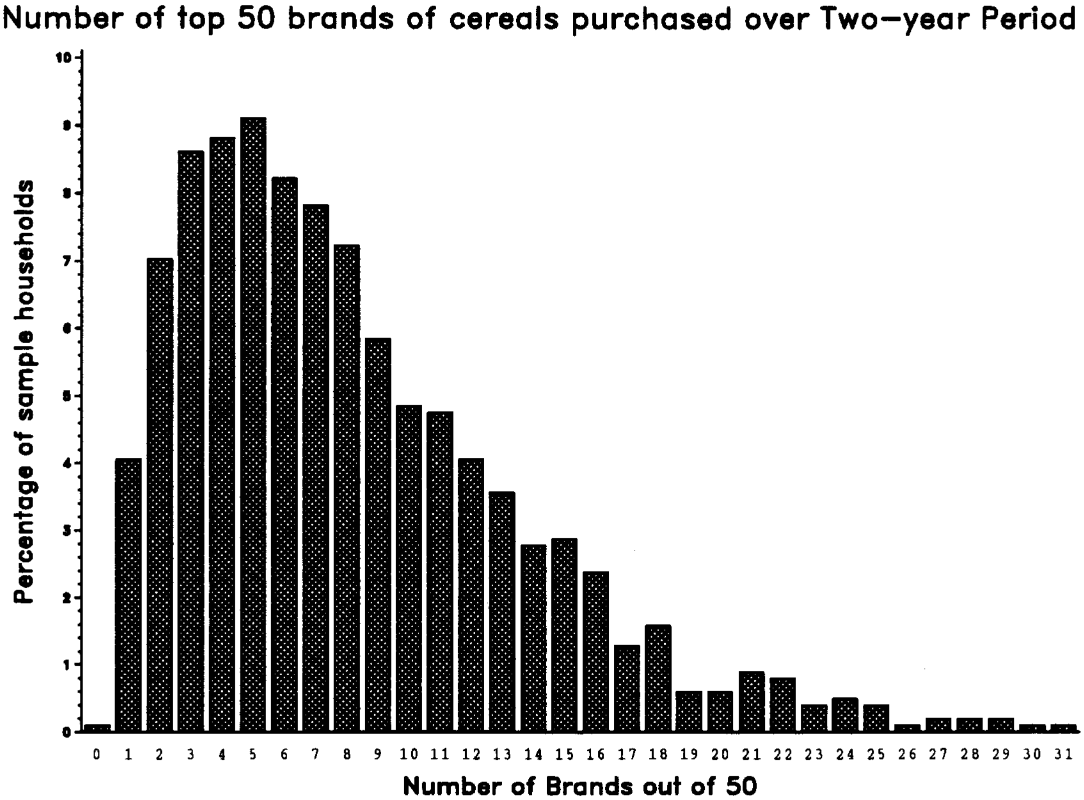


Figure 2.6: Number of top-fifty brands that a household consumes in 2-year sample period Average is 8.20 brands over 2-year sample period.

These summary statistics must be interpreted with caution, however. On the one hand, it is easy to find evidence of habit persistence in brand choice behavior here. On the other hand, such persistence could simply be a reflection of utility maximization: in fact, in a discrete-choice world without stochastic elements, consumers would choose the same brand (that which yields maximal utility) over and over. Distinguishing between habit persistence, which arises due to experience and persistence arising from utility maximization requires a choice model in which differences in brand utilities are adequately controlled for at the household level. Such a model is described in the next section.

Relation to recent empirical work My work ties together two different empirical literatures. Marketers have long focused their attention on measuring the effects of habit persistence (or brand loyalty) and advertising on consumers' purchase probabilities and price sensitivities (see the survey article by Kaul and Wittink [43]). In particular, recent papers by Ackerberg [2] and Deighton, et. al. [29] have focused on the interaction between advertising and experience. They also find that advertising's effects are significantly more muted once a consumer has actually tried a product.

However, these studies have mostly been limited to product categories with few brands; my focus on modeling households' choice among 51 brands of cereal has more in common with recent empirical IO studies of product differentiated industries (auto: Berry, Levinsohn, and Pakes [16], Goldberg [33]; pharmaceutical: Stern [68]). Despite the importance of advertising (and other forms of non-price competition[30]) in many of these industries, these studies have by and large focused on (short-run) price competition, abstracting away from the influences that non-price variables can have on the (long-run) competitive environment of a product-differentiated market. An exception is Bresnahan, Stern, and Trajtenberg's[18] recent analysis of the effect of a non-price variable — product innovation — on product differentiation in the personal computer market.

[30]Brand proliferation and research and development can also fall under the rubric of non-price competition. Roughly speaking, most forms of non-price competition fall into Sutton's [71] analytical framework as "endogenous sunk costs": investments whose scale is determined by the number of competitors in the market. In evaluating the importance of each type of non-price competition as a barrier to entry, the most relevant consideration, it seems, is whether it is more profitable (on the margin) for an incumbent or an entrant to increase the non-price variable.

2.2 Modeling Brand Choice: A Discrete-choice Model

The main empirical investigation in this chapter concerns testing whether advertising reinforces or undoes the effects of habit persistence on brand choices at the household level. More specifically, I want to measure how a household's probability of purchasing a given brand depends on its previous experience with that brand, and on that brand's advertising. This involves specifying a model of these probabilities which is (1) amenable for individual-level data (i.e., I am not estimating an aggregate market share model); (2) relatively parsimonious, given that each household has a probability of purchasing any of 51 brands on any purchase occasion.

Given these considerations, I decided to derive the expressions for the purchase probabilities from a discrete-choice model of household-level brand choice. A sizeable literature has stressed the random utility interpretation of the discrete choice model (see Anderson et. al. [4] for a survey). This interpretation appears inappropriate here since the discrete choice model abstracts away from households' continuous *quantity* choices, which are an integral component of the utility that households derive from their cereal purchases. However, Dubin and McFadden ([30], pg. 347) develop a discrete-continuous model in which agents choose the (discrete) alternative which yields maximal *indirect utility*, where an alternative's indirect utility presumes that an optimal quantity of the alternative is purchased. I follow their explication here.

On each purchase occasion, household h faces a choice of 52 (indexed $i = 0, \ldots, 51$)[31] mutually exclusive[32] alternatives: purchase one of the 51 brands of cereal (options $1, \ldots, 51$)

[31] the 51 cereal brands, plus the 0th nonpurchase option

[32] An alternative approach which would explicitly account for the multiple purchases phenomenon (and also for the quantity decision) would specify, at the household level (and perhaps time-aggregated to the quarterly level), demand functions for each of the fifty brands of cereals (a household-level version of the representative-household aggregate demand model in Hausman [37]). The main advantage of this approach would be the extra flexibility allowed in parameterizing the own- and cross-price elasticities of substitution between the competing brands. However, households purchase very few of the 50 brands, even when observations are aggregated up to the quarterly level. The large number of resulting "zeros" present huge theoretical and computational hurdles. Theoretically, these zeros can be accommodated in a "virtual price" framework, as in Wolak [74] — but the dimensions of integration involved would be quite large. Even a simpler selection mechanism could involve difficult multidimensional multivariate normal integrals. In addition, the aggregation involved would shrink the time dimension of the dataset, which is an unattractive option given my emphasis on experience and repeat purchases.

or purchase no cereal at all (option 0). *Given choice of cereal brand i*, household h derives indirect utility

$$U_{iht} = u(i, M_h, p_{it}, p_{0t}, X_i, X_h, \eta_{iht}), \qquad i = 1, \dots, 51, \tag{2.2.1}$$

where: M_h is household h's income,p_{it} is brand i's price on purchase occasion t, p_{0t} is a price index of non-cereal grocery products, X_i are brand i characteristics, X_h are household h characteristics, and η_{iht} parameterize unobserved factors specific to (i, h, t).[33]

Conditional on non-purchase of cereal (i.e., $i = 0$), household h derives indirect utility

$$U_{0ht} = u(0, M_h, p_{0t}, z_i, z_h, \eta_{0ht}). \tag{2.2.3}$$

In the parlance of the discrete-choice literature, the nonpurchase alternative is labeled the "outside good".

On purchase occasion t, then, household h chooses the alternative i ($i = 0, \dots, 51$) which provides the most conditional indirect utility, as specified in equations (2.2.1) and (2.2.3):

$$\max_{i \in [0, \dots, 51]} U_{iht}. \tag{2.2.4}$$

The probability that alternative i is chosen then depends on the joint distribution of $(\eta_{0ht}, \dots, \eta_{51ht})$:

$$\text{Prob}\left\{\eta_{0ht}, \dots, \eta_{51ht} : U_{iht} > U_{jht} \text{ for } j \neq i\right\}. \tag{2.2.5}$$

[33]Implicit in the definition of the indirect utility function is that, in the second stage, the quantity of brand i which household h consumes, quan$_{iht}$, can be derived from the conditional indirect utility function in (2.2.1) via Roy's Identity:

$$\text{quan}_{iht} = \frac{-\partial u(i, M_h, p_{it}, p_{0t}, z_i, z_h, \eta_{iht})/\partial p_{iht}}{\partial u(i, M_h, p_{it}, p_{0t}, z_i, z_h, \eta_{iht})/\partial M_h}, \qquad i = 1, \dots, 51. \tag{2.2.2}$$

Since these quantity equations involve the same parameters as the indirect utility functions in (2.2.1), it is possible, given distributional assumptions on the η's, to derive the joint density of observed brand choices and quantities. Maximum likelihood estimation incorporating the quantity observations will yield more efficient estimates of the parameters. That is not attempted here, however, due to the awkwardness of the distributional assumptions and functional forms required to obtain analytical forms for the choice probabilities and quantity equations (for example, see Dubin and McFadden [30], pp. 350-351). However, simulated method of moments procedures (e.g., McFadden [52]) eliminate the need to derive analytic expressions for the choice probabilities. The possibility of extending these methods to handle the discrete-continuous setting will be explored in future work.

The specific distribution assumptions on $(\eta_{0ht}, \ldots, \eta_{51ht})$ employed in this chapter will be discussed later.

Modeling differences in preferences between experienced and inexperienced users Given my focus on advertising, I want to be as flexible as possible in specifying household preferences, especially differences across households in their response to advertising. In the discrete choice setting, the most flexible specification would assign "random coefficients" for (thereby permitting preferences to be idiosyncratic to) each household and brand, Estimation of such a model would proceed by specifying a joint distribution for the random coefficients, and then "integrating out" over this joint distribution in deriving the estimating likelihood function.

In the interests of computational feasibility, I restrict myself to a *binomial* distribution over the random coefficients; in other words, I assume that on each purchase occasion t, the coefficients which describe household h preferences for brand i — call these Γ_{iht} — can take on one of two possible sets of values — call them Γ_{iht1} and Γ_{iht2}. The question then arises as to which households should be assigned Γ_{iht1}, and which one Γ_{iht2}.[34]

My focus on the advertising — experience tradeoff suggests that a natural starting point is to assign coefficients to households based on their experience with brand i: in particular, assign Γ_{iht1} to household h if it has purchased brand i recently, and Γ_{iht2} otherwise. This method has the advantage also of allowing the random coefficients to change over time, as the set of brands that a household is defined to have "recently purchased" changes.

In other words, I assume that the utility that household h derives from brand i ($i = 1, \ldots, 50$)[35] on purchase occasion t depends on whether or not household h has experienced brand i recently:

$$U_{iht} = \begin{cases} U_{iht1} \equiv V_{iht1} + \eta_{iht1} \equiv X'_{iht}\Gamma_{iht1} + \eta_{iht1} & \text{if household } h \text{ has experienced brand } i \\ U_{iht2} \equiv V_{iht2} + \eta_{iht2} \equiv X'_{iht}\Gamma_{iht2} + \eta_{iht2} & \text{otherwise} \end{cases} \tag{2.2.6}$$

[34] A binomial distribution for the random coefficients was also assumed in Berry, Carnall, and Spiller [14].

[35] As explained below in section 2.3.1 , the composite 51-st brand and the nonpurchase options ($i = 51$ and $i = 0$, respectively) are treated differently.

From the econometrician's point of view, both the experienced and inexperienced utilities (U_{iht1} and U_{iht2} respectively) are composed of a deterministic component (V) and an unobserved component (η). Households, however, observe $\{\eta_{iht1}, \eta_{iht2}\}, i = 0, \dots, 50$[36] when choosing a brand of cereal. Furthermore, note that in this specification, I allow experience to affect both the deterministic (V) and stochastic (η) portions of a household's utility. The former is accommodated via interactions of the experience variable with price and advertising, while the latter is accommodated in a nested logit model, as explained below. [37]

There are two problems associated with implementing such a model in which coefficients of the utility specification are allowed to differ depending on whether households have recently purchased the brand. Firstly, as pointed out in the large literature on state dependence and unobserved heterogeneity (see Heckman [38]), the effects of past purchase on current purchases could be estimated with bias in the presence of *unobserved heterogeneity*: unobserved propensities of a given household to purchase given brands can, when not controlled for, lead to "spurious" habit persistence. Secondly, as described in the previous section, there is some evidence that households in my sample are observed to purchase cereals much less frequently than households in the general US population. This may be due

[36]The composite 51st brand is treated slightly differently; see equation 2.3.20 and its discussion below.

[37]The model specified in equations 2.2.4 and 2.2.6 is *myopic*, because I assume that household h simply chooses the brand i which maximizes current utility U_{iht}, without considering the possibility that consuming brand i today will cause it (the household) to be habituated to it in the near future. A fully consistent dynamic treatment of the household's decision problem would involve choices of brands which maximize the expected value of the future stream of utilities. Ackerberg [1] and Erdem and Keane [31] have developed and estimated dynamic programming models of a household's brand choice decision allowing for households to gradually learn about the quality of different brands. However, both of these models involved a limited number of brands; accommodating the top fifty brands of cereal would lead to an intractable model, since the "state" variable in such a model would include not only a household's latest estimates of the quality of each brand, but also the prices, advertising expenditures, etc — all of the exogenous covariates — of each brand. Solving for the implicitly-defined value function would be computationally forbidding.

In a recent paper, Chintagunta, Kyriazidou, and Perktold [24] derive conditions under which a "myopic" discrete choice model (similar to the one specified above) can be rationalized as resulting from a forward-looking brand choice process. First of all, they assume that all the included covariates are i.i.d. over time; this assumption precludes dependent processes (such as time trends, or a learning model as in the papers cited in the previous paragraph). Furthermore, they model habit persistence as a series of indicators whether or not a brand was consumed on the previous purchase occasion. Since my definition of persistence differs from theirs (in my case persistence is an indicator of whether or not a given brand was consumed in a fixed period of time prior to the current purchase occasion), my model cannot be rationalized as being derived from a forward-looking decision process, even given the *iid* over time assumption on the included covariates.

to the omission of purchases at non-scanner stores. Discussion of unobserved heterogeneity will be delayed until a later section, while the probabilistic nested logit model which I develop to control for the possibility of unobserved purchases is described next.

2.2.1 Controlling for Unobserved Purchases: a Probabilistic Nested Logit (PNL) Model

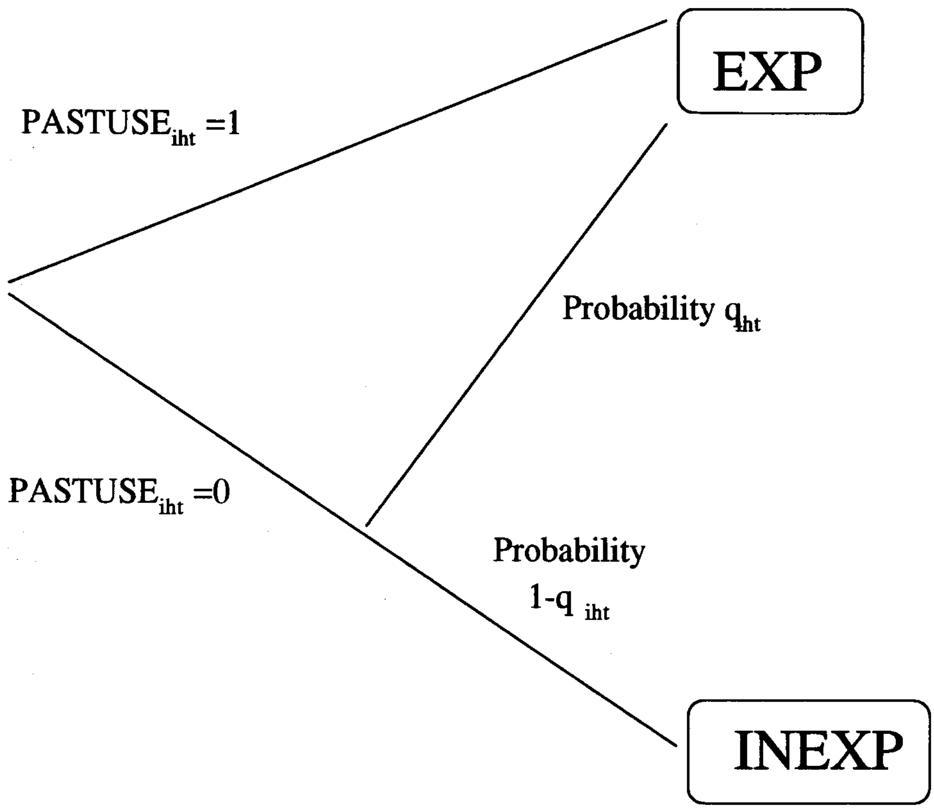
A household's recent experience of a brand is not always observed by the econometrician. To account for possible underestimation of households' experiences with brands due to unobserved purchases at non-scanner stores, I develop what I call an probabilistic nested logit (hereafter PNL)[38] model where I explicitly classify, for each given brand and each purchase occasion, each of the households in my sample as either *experienced* or *inexperienced* with the brand.

Refer to figure 2.7, which maps out the structure of the PNL model for a particular household h, brand i, and purchase occasion t. Clearly, if household h is observed to have purchased brand i recently (i.e., $\text{PASTUSE}_{iht} = 1$, corresponding to the upper branch in figure 2.7), I automatically categorize it as "experienced" (labeled EXP) with brand i on purchase occasion t. If household h is not observed to have purchased brand i recently (i.e., $\text{PASTUSE}_{iht} = 0$, the bottom branch in fig. 2.7), I estimate the probability q_{iht} that they are nevertheless experienced with the brands as a result of purchases which I do not observe in my dataset. q_{iht} will be parameterized in terms of (exogenous) brand- and household-specific covariates which are I expect to be correlated with a household's propensity to purchase a given brand.

While I motivate this model as one which explicitly controls for the possibility of unobserved purchases it is, however, more generally, a model which allows households to become *experienced* with a brand in ways other than directly purchasing the brand at the six stores included in the sample. Besides purchases at stores outside the six supermarkets, these other ways of becoming "experienced" could include trying out brands at restaurants or friends'

[38]This model shares some similarity with switching regression models with unobserved regime (i.e., unknown sample separation)

Figure 2.7: Structure of the Probabilistic Nested Logit (PNL) brand choice model
This diagram describes household h's preferences for brand i



EXP: Household h classified as experienced with brand i
INEXP: Household h classified as inexperienced with brand i

houses. Furthermore, if one subscribes to the view that experience yields information about the attributes of a brand, than the dichotomy between experienced and inexperienced brands can be reinterpreted as one between brands that a given household is "more" or "less" informed about. Then q_{iht} describes the probability that households become informed about the attributes of a product in ways other than through direct purchases.

No matter the interpretation, the important feature of the model is that the parameters in the indirect utility that a household derives from a brand differs depending on whether it is experienced with a brand. These differences in parameters (as well as partial observability of a household's experience status, for the observations where PASTUSE=1) serve to identify the parameters of the probability q_{iht}.

The utility specification For the econometrician, the uncertainty in the above utility specification (2.2.6) concerns whether or not a household is experienced with a brand j for which $\text{PASTUSE}_{jht} = 0$ (i.e., no recent purchases of this brand were observed). The econometrician uses the following index function to predict whether or not a household has made unobserved purchases of this brand:

$$y^*_{iht} = Z'_{iht}\Theta + \nu_{iht}. \tag{2.2.7}$$

This function summarizes, in reduced-form manner,[39] the econometrician's conception of household h's propensity to purchase brand i (more about this below). q_{iht} is then defined as $\text{Prob}(\nu_{iht} > -Z'_{iht}\Theta)$. From the econometrician's point of view, then:

$$U_{iht} = \begin{cases} V_{iht1} + \eta_{iht1} & \text{if PASTUSE}_{iht} = 1 \text{ or } y^*_{iht} > 0 \\ V_{iht2} + \eta_{iht2} & \text{otherwise} \end{cases} \tag{2.2.8}$$

for $i = 1, \ldots, 50$. At this point, I do not estimate unobserved purchase probabilities for the composite 51st brand. Discussion of the regressors which enter V and Z will be delayed to section 2.3.1.

[39]Further discussion in section 2.3.1.

Distributional assumptions Assuming independence over observations, the joint distribution of the η's and ν's for a given observation determines the form that the resulting brand choice probabilities will take. Introduce the notation that:

$$\eta^*_{iht} \equiv \begin{cases} \eta_{iht1} & \text{if household } h \text{ is experienced with brand } i \\ \eta_{iht2} & \text{otherwise.} \end{cases} \quad , \ i = 1, \ldots, 50 \qquad (2.2.9)$$

and define $\tilde{\nu}_{\mathbf{ht}} \equiv \{\nu_{iht} : \ \forall\, i \text{ s.t. } \mathrm{PASTUSE}_{iht} = 0\}$, the vector of disturbances corresponding to the brands. which household h is not observed to have bought recently. I make the following assumptions on the stochastic η^*'s and ν's:

- η^*'s and ν's are known to households, but unobserved by econometrician
- η^*'s are unobserved brand attributes, which vary over households, over time, and allowed to differ for a given household depending on whether it is experienced with brand i. η^*_{iht}'s are *i.i.d.* across h and t, but allowed to be correlated across i for a given (h,t).
- ν's are unobserved (to the econometrician) components describing the unobserved propensities of household h to purchase brand i. ν_{iht}'s are *i.i.d.* across i, h, and t. The *i.i.d.* across time assumption is particularly strong, given that by definition experience persists for 12 weeks. Implicitly any serial correlation in household h's unobserved propensity to purchase brand i at a non-scanner store at time t (parameterized by the latest variable y^*_{iht}) is only captured by the covariates Z_{iht}.[40]
- For a given (h,t), each of the η's are are independent of $\tilde{\nu}_{\mathbf{ht}}$. This implies that, conditional on the Z's, there is no correlation between y^*_{iht}, household h propensity to make "unobserved purchases" of brand i, and the η's. This is a strong assumption, and I plan to relax it in future work by allowing some arbitrary correlation between η^*_{iht} and ν_{iht}.[41]
- Each ν_{iht} is distributed standard normal.

Nested Logit Model with Probabilistic Nests One's choice of the joint distribution of $(\eta_{0ht}, \eta^*_{1ht}, \ldots, \eta^*_{50ht}, \eta_{51ht})$, conditional on the vector $\vec{\nu}_{ht} \equiv (\nu_{1ht}, \ldots, \nu_{50ht})$, determines the form that the resulting choice probabilities will take. Multinomial logit choice probabilities

[40] A future extension will allow for serial correlation over t in the ν_{iht}'s.

[41] One possibility is to let η^*_{iht} and ν_{iht} to be correlated via a common component, and then assume a discrete distribution for this common factor. This is discussed in Mroz and Guilkey [58].

result if I assume that each η^*_{iht} is distributed *i.i.d.* type II extreme value, conditional on the vector $\vec{\nu}_{ht}$.

As is well known, however, the multinomial logit model has the unattractive "Independence of Irrelevant Alternatives" property, which restricts any pair of brands to substitute equally well with a third brand.[42] Nested logit models, in which clusters of brands which are deemed by the researcher to be particularly "close" substitutes are grouped together into "nests", overcome this property, but their main drawback is a certain arbitrariness in dividing the alternatives into mutually exclusive nests. Therefore, I tried out several different specifications of the nesting structure.

I began by nesting on the basis of brand segment (i.e., FAMILY, ADULT, and KIDS nests), as well as nesting the nonpurchase option apart from all other alternatives, but these specifications yielded estimates of the substitution parameter σ (see below) which were outside the range in which household behavior could be rationalized by utility maximization of the kind assumed in equation 2.2.4. Thus after some trial and error, I settled on the following specification, which assumes that, at the household-level, experienced and inexperienced brands are in different nests. Note that this specifications implies that nests vary over households (since households differ in the brands that they are experienced with) and, for a given household, over time. This nesting structure is illustrated in figure 2.8.

More formally, assume that the vector $(\eta^*_{iht}, i = 1, \ldots, 50)$ is distributed generalized extreme value[43] (the h and t subscripts are omitted for clarity):

$$F(\eta_0, \eta^*_1, \ldots, \eta^*_{50}, \eta_{51} \mid \nu_{iht}, i = 1, \ldots, 50) = \exp\left[-G(e^{\eta_0}, e^{\eta^*_1}, \ldots, e^{\eta^*_{50}}, e^{\eta_{51}})\right]. \quad (2.2.10)$$

I assume that the G function above takes the form:

$$G(Y_0, \ldots, Y_{51}) = Y_0 + \left(\sum_{j\in 1} Y_j\right)^{1-\sigma} + \left(\sum_{j\in 2} Y_j\right)^{1-\sigma} + Y_{51} \quad (2.2.11)$$

and **1** denotes the set of brands that household h has experienced recently, and **2** denotes the set of brands that it has not recently (i.e., in the most recent 12 weeks) experienced.

[42] That is, the cross-price elasticities $\epsilon_{ik} = \epsilon_{jk}$, for any pair of brands $(i, j) \neq k$.

[43] The assumption of independence between the η^*'s and ν's is required here; if this were not the case, the conditional joint distribution of the η^*'s would most likely not generalized extreme value.

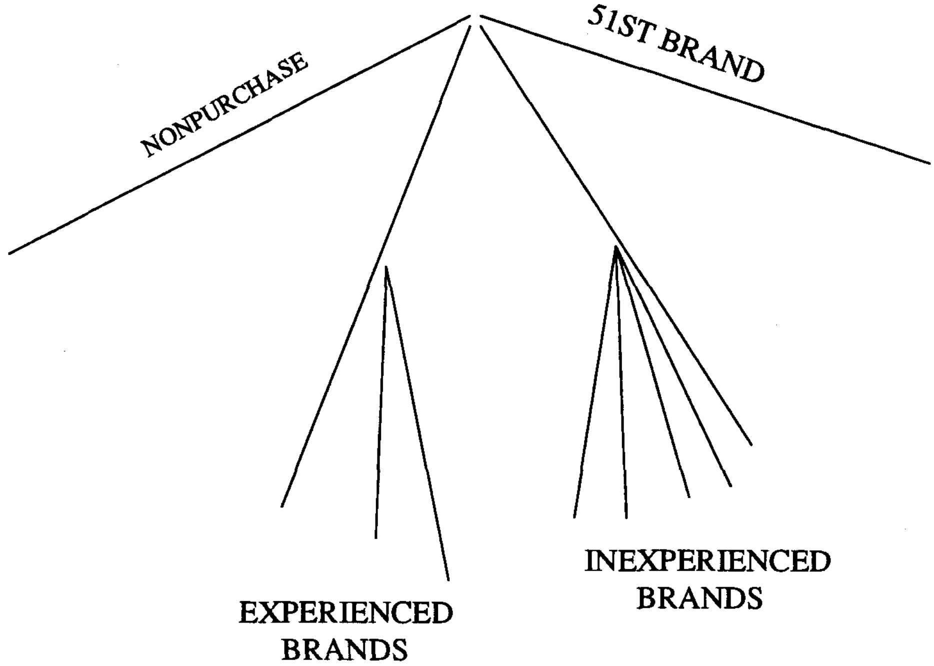


Figure 2.8: Nesting structure of the nested logit model

By allowing brands to be "nested" for a given household on the basis of experience, I essentially permit the stochastic utility component η to differ systematically across households depending on whether the household is experienced with the brand. This is clear from Cardell's [19] more intuitive *variance components* interpretation of the errors in a nested logit model: he views these error terms as composed of an "alternative-specific" component which is *i.i.d.* over all alternatives, as well as a "nest-specific" component which is fixed for all alternatives within a nest, but *i.i.d.* over the different nests in a given tier of the model (analogous to the time-invariant individual-specific "fixed effect" in panel data models). Furthermore, the σ parameter measures the relative importance of each component, with a larger σ indicating greater importance of the nest-specific component. In my model, then, σ can be taken as a measure of the importance of experience on the stochastic part of utility.

Probabilistic nests Since the econometrician doesn't fully observe which brands household h is experienced with on purchase occasion t, the likelihood function for this observation will involve integrating over the errors $\tilde{\nu}_{\mathbf{ht}}$.

For convenience, introduce the shorthand notation that

$$\mathrm{K}_{iht} = \mathrm{PASTUSE}_{iht} + (1 - \mathrm{PASTUSE}_{iht}) * \mathbf{1}(\nu_{iht} > -Z'_{iht}\Theta). \tag{2.2.12}$$

In other words, K_{iht} is an indicator for whether household h is experienced with brand i, whether by having made recent purchases of it in the dataset (i.e., $\mathrm{PASTUSE}_{iht} = 1$) or having a draw of ν_{iht} which exceeds the value of the index function $Z'_{iht}\Theta$.

From equations 2.2.10 and 2.2.11, then, the likelihood (*conditional on* $\tilde{\nu}_{\mathbf{ht}}$) that household h purchases brand i on purchase occasion t takes a nested-logit form:

$$d^*_{iht} \mid \tilde{\nu}_{\mathbf{ht}} = \frac{\begin{array}{c}(K_{iht} * \exp(V_{iht1}) + (1 - K_{iht}) * \exp(V_{iht2})) - \\ \left[\sum_{j=1}^{50} \mathbf{1}(K_{jht} = K_{iht})\,(K_{iht} * \exp(V_{iht1}) + (1 - K_{iht}) * \exp(V_{iht2}))\right]^{\sigma_h}\end{array}}{\exp(V_{0ht}) + \exp(V_{51ht}) + \left(\sum_{j=1}^{50} K_{jht} * \exp(V_{jht1})\right)^{1-\sigma_h} + \left(\sum_{j=1}^{50} (1 - K_{jht}) * \exp(V_{jht2})\right)^{1-\sigma_h}} \tag{2.2.13}$$

and similarly for purchases of the "outside good"

$$d^*_{0ht} \mid \tilde{\nu}_{\mathbf{ht}} = \frac{\exp(V_{0ht})}{\exp(V_{0ht}) + \exp(V_{51ht}) + \left(\sum_{j=1}^{50} K_{jht} * \exp(V_{jht1})\right)^{1-\sigma_h} + \left(\sum_{j=1}^{50} (1 - K_{jht}) * \exp(V_{jht2})\right)^{1-\sigma_h}} \tag{2.2.14}$$

as well as the 51st composite brand

$$d^*_{51ht} \mid \tilde{\nu}_{\mathbf{ht}} = \frac{\exp(V_{51ht})}{\exp(V_{0ht}) + \exp(V_{51ht}) + \left(\sum_{j=1}^{50} K_{jht} * \exp(V_{jht1})\right)^{1-\sigma_h} + \left(\sum_{j=1}^{50} (1 - K_{jht}) * \exp(V_{jht2})\right)^{1-\sigma_h}}. \tag{2.2.15}$$

Now, the unconditional likelihood function involves integrating out over a chosen joint distribution for $\tilde{\nu}_{\mathbf{ht}}$:

$$L(i, h, t) = \int (d^*_{iht} \mid \tilde{\nu}_{\mathbf{ht}}) dF(\tilde{\nu}_{\mathbf{ht}}), \qquad i = 0, 1, \ldots, 51 \tag{2.2.16}$$

Similarities with Probabilistic Choice Set (PCS) models This PNL model shares many similarities with the probabilistic choice set (hereafter PCS), or "consideration set" models employed in transportation economics and marketing (see Ben-Akiva and Boccara [10] and Andrew and Srinivasan [5]). These models explicitly accommodate the researcher's uncertainty about the size and composition of agents' choice sets by permitting them to be random. The index functions (2.2.7) now parameterize whether a household would have a particular alternative in its choice set: Swait [72] calls them "random constraints" in transportation mode choice models in which the researcher does not observe whether an agent has access to a particular mode of transportation.

The multinomial likelihood functions which arise in these PCS or consideration set models become intractable as the cardinality of the "universal" choice set (the set of all possible alternatives) grows large. Chib, et. al. [23] have estimated this model using Bayesian techniques which avoid direct maximization of the intractable likelihood function. The methodology I have proposed above overcomes this intractability, thereby permitting direct maximization of the likelihood function. I overcome the intractability of a multinomial likelihood function by specifying a continuous multivariate distribution for the disturbances (i.e., the ν's) in the index function (2.2.7) which induces a multinomial likelihood function when integrated over.[44] The integration is done using simulation techniques, as explained in the next section. This methodology is directly applicable to the PCS or consideration set models developed in the papers cited above.

[44]In other words, the unconditional likelihood function in equation 2.2.16 can be rewritten in the multinomial form

$$L(i,h,t) = \sum_{\rho in \mathcal{P}} \text{prob}_{ht}(\rho) * \text{prob}_{ht}(\text{alternative } i \mid \rho), \qquad \text{i=0, \dots ,51.} \tag{2.2.17}$$

$\rho \equiv (d_{\rho 1}, \dots, d_{\rho 50})$ specifies a particular partition of the top fifty brands into two mutually exclusive nests ($d_{\rho i}$ =1 if brand i is in the EXP nest and 0 otherwise, $i = 1, \dots, 50$). $\mathcal{P}$ is the set of all such possible partitions. prob(ρ) is the probability that this partition arises, which is equal to $\Pi_{i=1}^{50}\text{prob}_{ht}(d_{\rho i})$, where

$$\text{prob}_{ht}(d_{\rho i}) = \begin{cases} 1 & \text{if } d_{\rho i}\text{=1 and PASTUSE}_{iht}\text{=1} \\ 0 & \text{if } d_{\rho i}\text{=0 and PASTUSE}_{iht}\text{=1} \\ \text{prob}(\nu_{iht} > -Z'_{iht}\Theta) & \text{if } d_{\rho i}\text{=1 and PASTUSE}_{iht}\text{=0} \\ 1-\text{prob}(\nu_{iht} > -Z'_{iht}\Theta) & \text{if } d_{\rho i}\text{=0 and PASTUSE}_{iht}\text{=0} \end{cases} \tag{2.2.18}$$

These probabilities are induced by the distributional assumptions on the ν_{iht}'s. prob$_{ht}$(alternative $i \mid \rho$) is the choice probability of alternative i, given partition ρ; this is just the nested logit choice probability for alternative i with the composition of the two nests given by ρ.

2.3 Specification and Estimation Issues

2.3.1 Parameterizing the model

Recall from equation 2.2.6 that V_{iht}, the deterministic component of household h's indirect utility from consuming brand i on purchase occasion t, differs depending on whether household h is experienced (V_{1iht}) or not (V_{2iht}) with brand i on this purchase occasion:

$$\begin{aligned} V_{1iht} &= (X_i'\beta_0 + \delta_1) + (\alpha_1 + \delta_2) * p_{it} \\ &\quad + \alpha_2 * \mathrm{ADV}_{it} + \alpha_3 * \mathrm{ADV}_{it} * p_{it} + X_{ih}'\beta_1 + X_{2ht}'\beta_2 \\ V_{2iht} &= X_i'\beta_0 + \alpha_1 * p_{it} \\ &\quad + (\alpha_2 + \delta_3) * \mathrm{ADV}_{it} + (\alpha_3 + \delta_4) * \mathrm{ADV}_{it} * p_{it} + X_{ih}'\beta_1 + X_{2ht}'\beta_2 \end{aligned} \tag{2.3.19}$$

where:

- p_{it} is the price of brand i during purchase occasion t.
- X_i are characteristics which differ across brands. In the current specifications, I have just included brand dummies for each brand.
- X_{ih} are interactions between household demographics and brand characteristics which captures household-specific utility from brand i. The household-specific demographics are FAMSIZE and LOGINC, and the brand characteristics which I interact with the demographics are the segment dummies (FAMSEG, ADULTSEG, and, KIDSSEG), SUG, and FAT.[45]
- Since utility from cereal, as defined here, is specific to a short period of time — an inter-purchase occasion period — it will depend on factors specific to that particular purchase occasion. X_{2ht} contains regressors which capture these factors. Since cereals are a "stocked" item in most households, it seems natural to assume that, on such a short time scale, the utility that a household derives from a cereal purchase depends on the amount of cereal they have sitting around the house — clearly, the more they have sitting around, the less utility they will derive from an incremental cereal purchase. These "purchase frequency" dynamics are captured by the indicator variable PREVPURC, which is equal to 1 if a purchase of cereal occurred in the week previous to any purchase occasion.[46]

[45] A more general specification was tried in which I estimated separate coefficients on FAMSIZE and LOGINC for each brand — effectively "interacting household demographics with brand dummies" — but these parameters were not very precisely estimated, nor did the likelihood function increase by much, so I stopped working in this direction.

[46] Admittedly, an indicator of purchase in the previous week captures the dynamics of purchase frequency only rudimentarily; an attractive alternative would have been total *amount* of cereal purchased in the previous week. However, employing a binary variable such as PREVPURC facilitates simulating households' purchase

I allow for several differences in the indirect utility for a given brand depending on whether a household is experienced with it. First of all, I allow experience to shift households' utilities up or down through δ_1, and shift price sensitivities through δ_2. Secondly, I allow advertising's effect on utility to vary across experienced and inexperienced households, again in terms of shifts in utilities (δ_3) and price sensitivities (δ_4).

From a competitive point of view, the most important consideration is whether advertising's effects increase or decrease with experience. Assuming that experience leads to habituation, if advertising is more effective once a household has experienced the product, then advertising reinforces habit persistence ; this tends to reduce competition. On the other hand, advertising which is more effective on inexperienced households will encourage households to switch to less familiar brands, therefore promoting competition. Thus the competitive implications depend crucially on the empirical finding of whether advertising is more or less effective on inexperienced households — to borrow Shapiro's felicitous phrasing once again, whether advertising "substitutes" or "complements" experience. Roughly speaking: **in terms of the model parameters, advertising which substitutes for experience be captured by δ_3 and δ_4 having the same sign as δ_1 and δ_2, respectively.**

Similar hypotheses have also cropped up in consumer behavior research. In particular, the "usage dominance" hypothesis[47] states that once consumers experience a product, advertising's influence fades. In contrast, the "framing" or "validation" hypothesis posits that advertising's main role is to confirm or complete consumers' experience of a brand, essentially reassuring them that the appropriate choice was made. Such effects will actually increase with experience.

Comparing equation (2.3.19) to the conditional indirect utility function in equation (2.2.1), one notices several differences. First, p_{0t}, the price of the outside good, is missing from (2.3.19). As discussed below, I normalize $p_{0t} = 0$, $\forall\ t$. Second, income (y) enters

histories, which is done in order to calculate the long-term aggregate elasticities reported below. With a continuous variable, simulating purchase histories would also involve specifying a process for the *amount* purchased which, as discussed at the very beginning of this chapter, is not a component of the brand-choice model specified here.

[47] See the discussion in Deighton, et. al. [29] for more details.

(2.3.19) as LOGINC. Alternative, I could use total shopping expenditures during the week of purchase occasion t, but this variable could be correlated with unobserved factors specific to purchase occasion t which are captured in the η disturbances.

Household h's utility from the composite brand #51 during purchase occasion t, V_{51ht}, is parameterized as follows:

$$V_{51ht} = \alpha_{51} + (\alpha_1 + \lambda_1) * p_{51ht} + (\alpha_2 + \lambda_2) * \mathrm{ADV}_{51ht} + (\alpha_3 + \lambda_3) * p_{51ht} * \mathrm{ADV}_{51ht} + \lambda_4 * \log N_{51ht} \tag{2.3.20}$$

where p_{51ht} and ADV_{51ht} are, respectively, the store sales-weighted price and advertising averages for the brands included in the composite, $\mathrm{PASTUSE}_{51ht}$ is an indicator which takes on the value one if household h consumed *any* non-top fifty brand in the 12 weeks prior to purchase occasion t, and N_{51ht} is the number of brands included in the composite. This last covariate is included in the specification to capture the attractiveness of this composite brand due simply to the number of brands included (this covariate's function is discussed further in McFadden [51] and Ben-Akiva and Lerman [11] (pp. 254–260)). The parameters λ_1 to λ_4 are specific to the utility from the 51st brand. The first three of these parameters allow the coefficients α_1, α_2, and α_3 (respectively) to differ for this composite brand, since the covariates included here are averages, and therefore different from the covariates used for the other brands. The λ_4 parameter captures the inherent differentiation between the brands in this composite, and should lie in the unit interval.

Furthermore, note that experience (PASTUSE) does not enter the specification of V_{51} in equation 2.3.20: I do not permit experience to influence a household's choice of this 51st brand. I make this restriction because I want to abstract away from issues of how to define experience with a bundle of brands. If I were to define experience (i.e., PASTUSE) for this brand in the same way I define it for the top fifty brands — a household is experienced (i.e., $\mathrm{PASTUSE}_{51} = 1$) with brand #51 is it has tried *any* of the non top-fifty brands in the preceding 12 weeks — it is clear that being experienced with the 51st brand means something different than being experienced with one of the top fifty brands. I have estimated specifications allowing households to be experienced with this 51st brand in this manner, but also permitting experience's effects to differ for this brand. The results do not differ greatly from those reported below.[48]

The deterministic component of the indirect utility from the cereal nonpurchase option (the "outside good", in the terminology of the discrete-choice literature) has been normalized to zero: $V_{51ht} = 0, \ \forall \ t$. This implies that the price terms associated with the utility from the outside good has been subsumed into the constant terms in the utility specifications for each of the 51 brands considered here.

This can be considered a "normalization" in the cross-sectional case, but given panel data (as is the case here), by not permitting the constant terms for each of the 51 brands to vary over time, I am implicitly assuming that the "price" of the outside good — which includes all other commodities purchased during a shopping trip — stay constant over the sample period: $p_{0t} = 0, \ \forall \ t$.[49] I restrict σ_h, which parameterizes the larger substitutability within than across nests, to lie in the unit interval and to vary for households which contain

[48]However, results change noticeably when I permit experience to affect purchases of the 51st brand, but restrict its effects to be the same as for the top fifty brands — the "advertising substitutes for experience" result ($\delta_3 > 0$) becomes even stronger. Clearly, this is driven by the facts that purchases of the 51st brand constitute 25% of my observations (see the bottom line of figure 2.1) and that most of the brands aggregated into the 51st brands are less advertised. Given the large number of 51st brand purchases, most households will be classified as "experienced" with the 51st brand: the resulting high incidence of purchases of the often-experienced but low-advertised 51st brand will lead to a strong finding that experience and advertising are substitutes: granted, that's an interesting result from my point of view, but I don't wish for it to arise simply because of how I define experience with the 51st brand!

[49]One alternative would have been to explicitly deflate the prices for the 51 brands by (say) the nondurable CPI, but there is very little monthly variance in the CPI during the relatively short time-span of my data. From June 1991 to June 1993, the Chicago-area nondurables CPI rose about 7%; furthermore, most of the results below were estimated only on data from Jan–June 1992, when this CPI rose only about4% (**Source:** Bureau of Labor Statistics).

children (i.e.,, YGCHIL=1):[50]

$$\sigma_h = \frac{\exp(\sigma_1 + \sigma_2 * \mathrm{YGCHIL}_h)}{1 + \exp(\sigma_1 + \sigma_2 * \mathrm{YGCHIL}_h)}$$

$$\sigma_1,\ \sigma_2 \to -\infty \iff \sigma_h \to 0, \tag{2.3.21}$$

$$\sigma_1,\ \sigma_2 \to +\infty \iff \sigma_h \to 1.$$

Note that this reparameterization ensures that the σ's for *both* households with and without children lie in the unit interval.

Finally, the index function which determines q_{iht} represents, in reduced-form, the econometrician's conception of household h's propensity to have "experienced" brand i other than through direct purchases at the six scanner supermarkets. I use the following parameterization:

$$\begin{aligned} Z'_{iht}\Theta = {} & \theta_1 + \theta_2 * \mathrm{ADV}_{it} + \mathrm{YGCHIL}_h * (\theta_3 + \theta_4 * \mathrm{ADV}_{it}) + \\ & \mathrm{ADULTSEG}_i * \{\theta_5 + \theta_6 * \mathrm{ADV}_{it} + \mathrm{YGCHIL}_h * (\theta_7 + \theta_8 * \mathrm{ADV}_{it})\} + \\ & \mathrm{KIDSSEG}_i * \{\theta_9 + \theta_{10} * \mathrm{ADV}_{it} + \mathrm{YGCHIL}_h * (\theta_{11} + \theta_{12} * \mathrm{ADV}_{it})\}. \end{aligned} \tag{2.3.22}$$

This probability is parameterized as a function of brand characteristics (segment dummies and advertising) as well as household demographics (YGCHIL, which indicates whether or not a households has young children).

The reduced-form nature of this parameterization should be stressed. A structural approach to modeling past purchases would recognize that the probability that a household purchased a given brand in the past depends on the same parameters that determine its current probability of purchasing this brand. From the fundamentals of the brand-choice developed so far, It is possible (although difficult, since the purchase/nonpurchase decision much also be taken into account) to derive an explicit analytical formula for the probability of omitted purchases during the most recent 12 weeks. I do not do this for several reasons. First of all, any analytical formula places many nonlinear restrictions on the parameters, which may complicate estimation.

[50]While a negative σ_h would still yield valid choice probabilities (i.e., they are positive and would sum to 1), it is not consistent with choice probabilities which arise from the utility maximization problem in equation 2.2.4.

Secondly, as alluded to earlier in section 2.2.1, this model actually admits more flexible interpretations: for example, the distinction between experienced and inexperienced can be thought of as, more generally, one between brands that a household is "more" or "less" informed about. In fact, the structure of the PNL model resembles Grossman and Shapiro's [35] model of informative advertising, with the important difference that households are not restricted to have zero probability of purchasing brands that they are "uninformed" about (i.e., inexperienced with). The $Z'_{iht}\Theta$ index function then plays a role analogous to the "advertising technology" function in the Grossman and Shapiro model, which describes the "informativeness" of advertising.[51]

Price and Advertising Endogeneity A standard problem in aggregate discrete choice models (see Berry [13] for a theoretical discussion, and Stern [68] and Nevo [60] for empirical applications) is that price and advertising are correlated with some "quality" component ζ_i, persistent across households and over time, which is unobserved by the econometrician and thus subsumed into η_{ih} in (2.2.4). In disaggregate models (such as the one presented above), this problem doesn't necessarily go away: Kennan [44] notes that while *economically* individual households take price and advertising as exogenously given when making their discrete-choice decisions, there is still *econometric* endogeneity if the unobserved quality component is persistent enough across households. In this case firms set price and advertising as a function of ζ_i.

The problem of price endogeneity is mitigated to some extent in my data since I have heterogeneity in prices across households, and over time. In other words, the variation in prices for a given household over time has both an aggregate and a household-specific idiosyncratic component. To the extent that the idiosyncratic variation in prices is nontrivial, Kennan argues that any bias resulting from possible price endogeneity disappears in large enough samples.

[51]The importance of advertising in the Grossman-Shapiro model is also why I chose to feature advertising so prominently in the index function (2.3.22). in any case, alternative specifications without advertising in the index function yield virtually identical results.

However, my advertising variable enjoys no such idiosyncratic across-household variation. In the specifications below, therefore, I control for advertising endogeneity which arises due to an unobserved ζ_i which is assumed constant across households and over time. In some of the specifications, I include a complete set of 51 brand dummies, effectively "estimating" the ζ's as fixed effects.

2.3.2 Estimating the model

The dimension of the integral in equation 2.2.16 is $R_{ht} \equiv \sum_{i=1}^{50}(1 - \text{PASTUSE}_{iht})$, the number of brands which household h is *not* observed to have purchased recently on purchase occasion t. R_{ht} is often greater than 40, so that even given the assumption that the ν's are distributed *i.i.d.* standard normal, the multidimensional integral in equation 2.2.16 is not analytically intractable, but can be easily evaluated by Monte-Carlo simulation.

This is a two step procedure. First of all, for observation (h, t), draw M R_{ht}-dimensional vectors standard normal random variables: $\tilde{\nu}_{ht}^m \equiv (\nu_{1h}^m, \ldots, \nu_{R_{ht}h}^m)$, for $m = 1 \ldots M$.

Secondly, evaluate the likelihood function at each of the drawn vectors $\tilde{\nu}^m$'s. Analogously to (2.2.12), define

$$\mathrm{K}_{iht}^m = \text{PASTUSE}_{iht} + (1 - \text{PASTUSE}_{iht}) * \mathbf{1}(\nu_{iht}^m > -Z_{iht}'\Theta), \tag{2.3.23}$$

an indicator which is equal to 1 if either $\text{PASTUSE}_{iht} = 1$ or the *drawn* ν_{iht}^m exceeds the index function $Z_{iht}'\Theta$.

For a purchase of brand i, the simulated likelihood function is then:

$$L_{sim}(i, h, t) =$$

$$\frac{1}{M}\sum_{m=1}^{M} \frac{(K_{iht}^m * \exp(V_{iht1}) + (1 - K_{iht}^m) * \exp(V_{iht2})) - \left[\sum_{j=1}^{50} \mathbf{1}(K_{jht}^m = K_{iht}^m)\,(K_{iht}^m * \exp(V_{iht1}) + (1 - K_{iht}^m) * \exp(V_{iht2}))\right]^{\sigma_h}}{\exp(V_{0ht}) + \exp(V_{51ht}) + \left[\sum_{j=1}^{50} K_{jht}^m * \exp(V_{jht1})\right]^{1-\sigma_h} + \left[\sum_{j=1}^{50} (1 - K_{jht}^m) * \exp(V_{jht2})\right]^{1-\sigma_h}}. \tag{2.3.24}$$

Analogous expressions can be derived for the likelihood functions relating to purchase of the 51st brand as well as nonpurchase.[52] Due to the computational complexity of the

[52]The indicator functions in the simulated likelihood function (which enter through (2.3.23)) introduce

estimation procedure, I used M=5 to derive the results below. Larger values of M (10, 20) yielded similar results, but took much longer to estimate.

2.4 Controlling for Unobserved Heterogeneity

Any effects of experience measured using the foregoing model are valid only to the extent that the coefficients which differ between the experienced and inexperienced utility specifications (the δ's) accurately capture the "pure" effect of experience — i.e., changes in preferences arising solely because a household has experienced the brand. Heckman [38] has labeled this phenomenon *state dependence*, equivalent to what I've called habit persistence.

However, we have assumed before that households are utility maximizers who try brands that they "like" rather than sample randomly from the available brands. Utility maximization, therefore, also leads to apparently habit persistent behavior (or *spurious state dependence*, in Heckman's words) because households will be observed to consume the same brands — those which yield highest utility — over and over. Since spurious state dependence can exaggerate the extent of habit persistence, I need to disentangle pure from spurious state dependence by adequately controlling for time-invariant, household- and brand-specific heterogeneity which could lead to spurious state dependence.

The length of my panel tends to mitigate the effect of spurious state dependence arising from *unobserved* heterogeneity. A long panel increases the likelihood that the variation in experience across households and brands is due to time-varying variation in "runs" (consecutive purchases of the same brand) rather than time-invariant differences in purchase propensities. [53]

---

sharp nonlinearities. In the actual estimation process, I will smooth over these nonlinearities with a Gaussian (normal) kernel:

$$\mathbf{1}(\nu^{m}_{i'h} > -Z'\Theta) = \Phi(\frac{\nu^{m}_{i'h} + Z'\Theta}{s})$$

Equality holds as the bandwidth $s \to 0$. In the results below, I used s=0.01.

[53] The length of the panel appears crucial to distinguishing between pure and spurious state dependence. Chamberlain [22] (pg. 1279) notes that with a panel of less than four periods, it is impossible to make this distinction for the simplest binary logit model, since with such a short panel we cannot simultaneously identify the state dependence parameter and control for the agent-specific fixed effects.

Nevertheless, these unobservables remain "omitted variables" in the current specification, and as such will bias the coefficient estimates independently of spurious state dependence. In this section I describe my approach in controlling for two general types of unobserved cereal preference heterogeneity: either

1. households differ in their natural preference for cereal, or

2. households differ in their natural preference for particular brands of cereal.

I assume heterogeneity takes the following general form: there are parameters ω_{ih}, $\forall h$, $i = 1, \dots, 51$ which capture time-invariant unobserved factors which influences household h's propensity to consume brand i. Assume that the ω's enter U_{iht}, household h's indirect utility from brand i on purchase occasion t, directly, and are constant over purchase occasions[54]:

$$U_{iht} \equiv V_{iht} + \eta_{iht} = X_{iht}\Gamma_{iht} + \omega_{ih} + \eta_{iht}$$

where $X_{iht}\beta$ parameterizes observed heterogeneity across households in preferences for brand i.

Random effects approach: assumptions A "random effects" approach is taken, i.e., I assume that the vector $\Omega_h \equiv (\omega_{1h}, \dots, \omega_{51h})'$ is distributed *i.i.d.* over all households h, and orthogonal to the covariates X. Then the likelihood function corresponding to each household's purchase histories is integrated over a specified distribution of Ω.

The presence of lagged dependent variables — in this case PASTUSE$_{iht}$, as well as the lagged cereal purchase indicator PREVPURC$_t$ — among the covariates, complicates the procedure. To see this, let's abstract away from unobserved purchases. While PREVPURC is a function of only y_{t-1} — namely, $\mathbf{1}(y_{t-1} \neq 0)$ — recall that PASTUSE$_{iht}$ is a function of $(y_{t-1}, \dots, y_{t-12})$. Conditional on Ω_h and $(y_1, \dots, y_{12})$, the likelihood function for the

[54]If the ω's varied over purchase occasion, they would be indistinguishable from the other error term η_{iht}.

observed sequence $(y_{13}, \dots, y_T)$ is therefore

$$\begin{aligned} & f(y_T, \dots, y_{13} \mid \Omega_h, y_{12}, \dots, y_1) \\ = \quad & \textstyle\prod_{t=13}^{T} g(y_t \mid \Omega_h, y_{t-1}, \dots, y_1) \\ = \quad & \textstyle\prod_{t=13}^{T} g(y_t \mid \Omega_h, y_{t-1}, \dots, y_{t-12}) \end{aligned} \tag{2.4.25}$$

where the last equality arises by assumption of the model (experience is "forgotten" after 12 weeks).

A general random effects specification would involve specifying the joint distribution of $(\Omega, y_1, \dots, y_{12})$, perhaps making it conditional on observable household characteristics Z_h and parameters Λ — call this $F(\Omega, y_1, \dots, y_{12} \mid \mathbf{Z}_h, \Lambda)$ — and then integrating out with respect to this distribution to derive the likelihood contribution for each household h:

$$\log L_h = \int \prod_{t=13}^{T} g(y_t \mid \Omega_h, y_{t-1}, \dots, y_{t-12}) dF(\Omega, y_1, \dots, y_{12} \mid \mathbf{Z}_h, \Lambda). \tag{2.4.26}$$

Several computational problems arise in doing this. First of all, the dimension of the integral is very large, but that can be handled via simulation methods. What ultimately hampers implementation of (2.4.26) is specifying the joint distribution for the discrete variables $y_1, \dots, y_{12}$, each of which can take on one of 52 values. Any resulting joint distribution of $(y_1, \dots, y_{12})$ will be *multinomial*, and it is difficult to integrate out over multinomial distributions. Nonetheless, estimation using the likelihood function 2.4.26 constitutes a challenging extension.[55]

For the current set of results, however, I make the stronger assumptions that (i) the initial values $(y_1, \dots, y_{12})$ are exogenous scalars and (ii) $F_h(\Omega \mid y_1, \dots, y_{12})$ doesn't depend on the values of $(y_1, \dots, y_{12})$. *Essentially, these assumptions ensure that the distribution of the unobserved heterogeneity parameters Ω is invariant to the initial values of* PASTUSE_{iht}, $i = 1, 51$ *and* PREVPURC_t. Given these assumptions, the resulting random

[55] Furthermore, the methodology I develop in the appendix handles the exact same problem of "integrating out" over a multinomial distribution using simulation methods, a problem that has vexed researchers estimating "consideration set" models often used in marketing and transportation research (cf. Andrews and Srinivasan [5], Ben-Akiva and Boccara [10]). Adapting that framework to handle the integral in equation 2.4.26 appears straightforward.

effects logit log likelihood function involves integrating out over the joint distribution of Ω_{ih} for all the observations pertaining to household h:

$$L = \sum_h \log \left\{ \int \left[\prod_t^{T_h} \prod_{i=0}^{51} y_{iht} d_{iht} \mid \Omega_h \right] dF_h(\Omega) \right\} \tag{2.4.27}$$

where y_{iht} indicates whether household h bought brand i at time t, and the integral goes over the 51-dimensional support of Ω. The next section presents a simple parameterization of $F_h(\cdot)$ which is used in the estimations reported below.

2.4.1 Discrete distribution: "cereal lovers" vs. "cereal haters"

To simplify the integral in the likelihood function 2.4.27, I assume that each ω_{ih} is mean zero and has a two-point discrete distribution, i.e.,

$$\omega_{ih} = \begin{cases} \omega_{ih}^H & \text{with probability } r_{ih} \\ \omega_{ih}^L & \text{with probability } 1 - r_{ih} \end{cases} \tag{2.4.28}$$

I restrict r_{ih} and $(\omega_{1h}, \cdots, \omega_{51h})$ to be the same for all families with children, and the same for all families without children.

Secondly, I parameterize

$$\omega_{ih} = \gamma_i * \omega_h \tag{2.4.29}$$

where

$$\omega_h = \begin{cases} \omega_h^H & \text{with probability } r_h \\ \omega_h^L & \text{with probability } 1 - r_h. \end{cases} \tag{2.4.30}$$

Households differ fundamentally in ω_h, their unobserved liking for cereal. It may be useful to think of households whose tastes are characterized by ω_H as "cereal lovers", while those with tastes characterized by ω_L are "cereal haters". Given this liking however, their unobserved preference for brand i, ω_{ih} is just ω_h multiplied by a factor γ_i which parameterizes the degree that households who "like" (or dislike) cereal will like brand i. This is also a convenient way to allow unobservables to be correlated across brands, so that the ω_{ih} are not independent across i.[56]

γ_i is normalized to 1 for the composite brand 51. In addition, it is restricted to take on only three different values, corresponding to the three cereal segments (family, adult, and kids): γ_{FAM}, γ_{ADULT}, γ_{KIDS}.

For each given household, then, the distribution of unobserved heterogeneity is *binomial.* With probability r_{ih}, that household is a "cereal lover" whose unobserved preferences are described by (ω_h^H, γ_{FAM}, γ_{ADULT}, γ_{KIDS}). With probability $(1 - r_{ih})$, these preferences are described by (ω_h^L, γ_{FAM}, γ_{ADULT}, γ_{KIDS}). The resulting likelihood function for this household is then:

$$L_h = \log\left\{ r_h * \left[\prod_t^{T_h}\prod_{i=0}^{51} y_{iht}d_{iht} \mid \omega_h^H\right] + (1-r_h) * \left[\prod_t^{T_h}\prod_{i=0}^{51} y_{iht}d_{iht} \mid \omega_h^L\right]\right\} \tag{2.4.31}$$

The fundamental intuition in this modeling alternative is that households differ in their inherent "liking" of cereal; their unobserved preferences for the individual brands depend (solely) on this inherent liking for cereal.[57]

2.5 Estimation results

Due to the computational complexity of the PNL model, most of the results below were obtained using only one half-year's worth of data (about a fourth of the total observations). Only the 22,146 observations from Jan-June 1992 were used. The summary statistics are similar across all half-years in the sample.[58] Results from five specifications are presented below, in figures 2.9 and 2.10. These specifications differ in several ways.

Model A is the baseline model, which is a nested logit model (where experienced and inexperienced brands are grouped into separate nests for each observation) in which neither

[56]In more technical terms, this correlation between unobservables relaxes the Independence of Irrelevant Alternatives (IIA) property of the multinomial logit model, which restricts (unrealistically) that the cross-price elasticities ϵ_{ik} and ϵ_{jk} are equal, regardless of any large differences between brands i and j.

[57]I have estimated more complex specifications of unobserved heterogeneity, allowing for explicit differences in the unobserved heterogeneity parameters across all brands as well as across all households. The resulting model, which required simulation estimation methods, did not produce qualitatively different results. Results, not reported here, are available from the author.

[58]The models involving simulation were estimated on a Silicon Graphics 16 processor machine, utilizing parallel processing which significantly reduced the computational time.

Figure 2.9: Model estimates: *without unobserved heterogeneity*
Using average transaction prices

	A: Baseline NL model		B: PNL	
Parameter	*Estimate*	*StdErr*	*Estimate*	*StdErr*
α_1: Price	-1.571	0.063	-2.725	0.478
α_2: Adv[a]	0.006	0.045	0.006	0.102
α_3: Price*Ad	-0.101	0.014	-0.065	0.034
δ_1: PASTUSE*constant	-0.003	0.211	-0.002	0.157
δ_2: PASTUSE*price	1.265	0.073	2.526	0.401
δ_3: (1-PASTUSE)*adv	-0.267	0.041	0.002	0.106
δ_4: (1-PASTUSE)*price*adv	0.161	0.016	0.106	0.097
Prevpurc	0.395	0.036	0.383	0.098
Nesting params:[b]				
σ_1	-3.047	0.693	-5.004	18.307
σ_2	1.965	0.615	0.187	14.891
$\rightarrow \sigma$ (households w/o chil)	0.045		0.007	
$\rightarrow \sigma$ (households w/ chil)	0.297		0.008	
Brand dummies	yes		yes	
Interactions w/LogInc[c]	yes		yes	
Interactions w/Famsize[d]	yes		yes	
Estimate $\lambda_1 - \lambda_4$[e]	yes		yes	
Controlling for measurement error in past purchases	no		yes	
LogL:	-35092.28		-35620.58	
NumObs:	22146		22146	
M (# sim'd draws)	—		10	

[a] $ mills, quarterly expenditure
[b] See equation 2.3.21.
[c] LogInc interacted with price, advertising, segment dummies and nutritional characteristics
[d] Famsize interacted with price, advertising, segment dummies and nutritional characteristics
[e] These are parameters specific to the utility from the 51st, composite brand. See equation (2.3.20).

Figure 2.10: Model estimates: *with unobserved heterogeneity*
Using average transaction prices

	C: NL		D: NL w/o brand dummies		E: PNL	
Variable	*Estimate*	*StdErr*	*Estimate*	*StdErr*	*Estimate*	*StdErr*
α_1: Price	-1.698	0.055	-1.021	0.037	-2.018	0.684
α_2: Adv[a]	0.007	0.051	0.008	0.008	0.006	0.068
α_3: Price*Ad	-0.110	0.016	-0.037	0.006	-0.094	0.066
δ_1: PASTUSE*constant	-0.001	0.199	-0.0001	0.0014	-0.001	0.023
δ_2: PASTUSE*price	1.179	0.067	1.004	0.017	1.605	0.451
δ_3: (1-PASTUSE)*adv	-0.280	0.040	0.037	0.016	-0.252	0.200
δ_4: (1-PASTUSE)*price*adv	0.165	0.016	0.015	0.008	0.192	0.133
Prevpurc	-0.202	0.037	-0.149	0.034	-0.177	0.249
Nesting params:[b]						
σ_1	-0.533	0.085	-0.159	0.068	-1.019	1.373
σ_2	0.269	0.083	0.297	0.064	0.098	0.747
$\rightarrow \sigma$ (households w/o chil)	0.370		0.460		0.265	
$\rightarrow \sigma$ (households w/ chil)	0.434		0.535		0.285	
Unobs'd heterogeneity params:						
q (families w/o young chil.)	0.365	0.041	0.299	0.042	0.341	0.311
q (families w/ young chil.)	0.318	0.053	0.278	0.050	0.403	0.348
ω_h (families w/o young chil.)	1.146	0.049	1.246	0.048	1.190	0.341
ω_h (families w/ young chil.)	1.233	0.061	1.359	0.060	1.030	0.373
γ_{FAM}	1.614	0.079	1.941	0.092	1.344	0.742
γ_{ADULT}	1.836	0.083	2.167	0.097	1.493	0.627
γ_{KIDS}	1.807	0.090	2.139	0.102	1.666	0.694
Brand dummies	yes		no		yes	
Interactions w/LogInc[c]	yes		yes		yes	
Interactions w/Famsize[d]	yes		yes		yes	
Estimate $\lambda_1 - \lambda_4$[e]	yes		yes		yes	
Controlling for measurement error in past purchases	no		no		yes	
LogL:	-34079.51		-34327.32		-33892.22	
NumObs:	22146		22146		22146	
M (# sim'd draws)	—		—		10	

[a]$ mills, quarterly expenditure
[b]See equation 2.3.21.
[c]LogInc interacted with price, advertising, segment dummies and nutritional characteristics
[d]Famsize interacted with price, advertising, segment dummies and nutritional characteristics
[e]These are parameters specific to the utility from the 51st, composite brand. See equation (2.3.20).

measurement error in past purchases nor unobserved heterogeneity is controlled for. **Model B** controls for possible omission of purchases in the PNL framework. **Model C** and **Model E** are equivalent to models A and B, respectively, except that they control for unobserved heterogeneity. **Model D** is Model C, except instead of brand dummies for each brand I have entered brand-level covariates (such as segment dummies and nutritional content) to soak up differences across brands. All of these models are estimated using the *average transactions prices* described earlier. The main qualitative results obtained in this analysis already emerge in the baseline model (model A) results, and are robust even after controlling for systematic underestimation of past purchases and unobserved heterogeneity. Therefore, unless otherwise noted, the point estimates cited below all pertain to the model A results.

Habit persistence: evidence First of all, the price coefficient (α_1) is negative (point estimate -1.571), as expected, indicating a positive marginal utility of income. Secondly, the positive estimate for δ_2 (1.265) indicates that households are much less price sensitive with respect to brands that they have recently experienced. This is strong evidence that *households' brand choices exhibit strong habit persistence*: households become habituated in the sense of having a lower sensitivity towards price changes in brands that they have recent experienced. Habituation, in turn, implies sharper differentiation between experienced and inexperienced brands, because a larger rise in the price of the brand that it is habituated with is required for a given household to switch to a competing, inexperienced brand.

Advertising's asymmetric effects Advertising has asymmetric effects on price sensitivities. First, the positive and large (in magnitude) estimate for δ_4 (0.165) indicates that, similarly to experience, advertising lowers households' price sensitivities, but only for households who have not experienced the product. This echoes earlier findings reported in Deighton, et. al. [29] and Ackerberg [1]. Therefore, we confirm Shapiro's [66] argument that *advertising substitutes for (or "mimics") experience* in the sense that it duplicates the effects of experience (i.e., lowers price sensitivities) on households who have not actually experienced the brand.

For households who have experienced the brand, the negative estimate of α_3 (-0.101) indicates that advertising actually *increases* the price sensitivities of households who have recently purchased a given brand. This result implies that, relative to the pre-advertising situation, a lower rise in the price of the experienced brand suffices to persuade a household to switch to an inexperienced brand, thereby *reducing* the degree of differentiation between experienced and inexperienced brands.

Furthermore, this finding the advertising increases the price sensitivities of households towards brands they have recently experienced echoes one of the main findings in Kanetkar, Weinberg, and Weiss [42]: that households become *more* price sensitive towards brands (of aluminum foil and dry dog food) whose television ads they were exposed to more often. It also confirms Steiner's [67] rather informal assertions that not only are consumers are more price sensitive vis-a-vis higher-advertising, more "visible" products, but also retailers are readier to cut prices on highly-advertised items, in order to draw consumers into their stores.[59]

Furthermore, this finding that advertising affects the price sensitivities of households differently depending on a household's recent experience of the product provides a possible resolution to the conflicting results surveyed by Kaul and Wittink [43] regarding the effect of advertising on price sensitivity. In their survey of the empirical marketing literature, they found some disagreement as to whether *non-price* advertising raised or lowered consumers' price sensitivities. Finally, if we define advertising as "persuasive" if it increases purchase probabilities, then this finding indicates that advertising only persuades inexperienced households.

Robustness of advertising's estimated effects The negative estimate of δ_3 indicates, however, that once advertising's effects on price sensitivities are controlled for, the net effect of a brand's advertising on its purchase probability is *negative* (the point estimate is -0.252 for model E). Part of the reason for this is that much of the effects of advertising differences

[59]The retailer dimension is not covered in my present analysis, but this finding indicates that maybe I should incorporate more retailer-related regressors (such as store promotions, availability of store coupons, etc.) in future work.

across brands is already subsumed into the brand-level fixed effects. This is evidenced by comparing the results of model D, which does not include brand dummies. Here the estimate for δ_3 becomes positive (0.037) while α_3 and δ_4 have the same sign as before.

Recall that the brand fixed effects were included to control for brand-specific "unobserved quality" which could lead to price and advertising endogeneity. Note that once I remove the brand dummies, as in the Model D specification, the price coefficient for experienced households ($=\alpha_1 + \delta_2$=-1.021+1.004) is very close to zero, implying unrealistically low price sensitivities. Furthermore, note that the likelihood function value is substantially higher for the model with than without brand fixed effects. Thus, while the model with brand dummies is a statistically superior model of brand choice, comparison with the results from models without brand dummies indicates that the brand dummies soak up much of the variation in advertising in the data, leading to the negative estimate of δ_3. Thus given the data constraint (the availability of only data on aggregate brand-level advertising expenditures) it appears difficult to identify the level effects of advertising on the choice probabilities separately from time-invariant brand-specific fixed effects.[60]

Given these identification problems, I estimated several variants of Model E (results reported in figure 2.11) in order to gauge the robustness of the estimates of advertising's effects (notably the findings that $\alpha_3 < 0$ and $\delta_4 > 0$) presented earlier. Expecting that these identification problems would become less severe as the model is estimated using more data (due to more across-time variation in advertising), I re-estimated model E using a full year's worth of data (44,285 observations) — this is **model G** in figure 2.11. However, the coefficient estimates hardly change. Neither do they change perceptibly when I define PASTUSE as whether a particular brand has been tried in the previous *four* (as opposed to 12) weeks (**Model F**), or when I estimate the model using shelf prices (**Model H**).

Furthermore, I also estimated other variants of the PML model, results from which are not presented in the tables. I introduced cross-sectional variation into the advertising variable by imposing restrictions (such as: households without children are not exposed to kids

[60]This problem would appear to be present also in the aggregate discrete choice models estimated by Stern [68] and Nevo [60], in which similar aggregate advertising data is also used.

Figure 2.11: Additional PNL model estimates

Model F: Same as Model E except 4-week definition of experience ($\text{PASTUSE}_{iht} = 1$ if household h tried brand i during *four* weeks prior to time t

Model G: Same as Model E except estimated with data for one year (Jan.-Dec. 1992)

Model H: Same as Model E except estimated using *shelf prices*

	F:		G:		H:	
Variable	*Estimate*	*StdErr*	*Estimate*	*StdErr*	*Estimate*	*StdErr*
α_1: Price	-1.821	0.953	-2.000	0.352	-2.186	0.847
α_2: Adv[a]	0.006	0.418	-0.001	0.043	0.008	0.529
α_3: Price*Ad	-0.092	0.162	-0.062	0.036	-0.073	0.233
δ_1: PASTUSE*constant	-0.001	0.158	0.002	0.207	-0.001	0.116
δ_2: PASTUSE*price	1.489	0.614	1.577	0.247	1.381	0.547
δ_3: (1-PASTUSE)*adv	-0.161	0.264	-0.291	0.167	-0.307	0.454
δ_4: (1-PASTUSE)*price*adv	0.166	0.181	0.192	0.084	0.182	0.240
Prevpurc	-0.197	0.384	-0.096	0.183	-0.133	0.438
Nesting params:[b]						
σ_1	-1.298	2.726	-1.164	0.795	-0.092	1.285
σ_2	-0.104	1.126	-0.588	1.202	-0.271	0.657
$\rightarrow \sigma$ (households w/o chil)	0.215		0.238		0.477	
$\rightarrow \sigma$ (households w/ chil)	0.198		0.148		0.410	
Unobs'd heterogeneity params:						
q (families w/o young chil.)	0.379	0.419	0.422	0.359	0.311	0.542
q (families w/ young chil.)	0.358	0.548	0.394	0.502	0.345	0.534
ω_h (families w/o young chil.)	1.546	0.469	1.061	0.394	1.249	0.580
ω_h (families w/ young chil.)	1.502	0.861	0.976	0.484	1.057	0.500
γ_{FAM}	0.910	0.645	1.152	0.409	1.927	1.199
γ_{ADULT}	1.073	0.664	1.345	0.389	2.121	1.160
γ_{KIDS}	1.183	0.764	1.440	0.704	2.351	1.412
Brand dummies	yes		yes		yes	
Interactions w/LogInc[c]	yes		yes		yes	
Interactions w/Famsize[d]	yes		yes		yes	
Estimate $\lambda_1 - \lambda_4$[e]	yes		yes		yes	
Controlling for measurement error in past purchases	yes		yes		yes	
LogL:	-35460.61		-65774.74		-33772.45	
NumObs:	22146		42185		22146	
M (# sim'd draws)	10		10		10	

[a]$ mills, quarterly expenditure

[b]See equation 2.3.21.

[c]LogInc interacted with price, advertising, segment dummies and nutritional characteristics

[d]Famsize interacted with price, advertising, segment dummies and nutritional characteristics

[e]These are parameters specific to the utility from the 51st, composite brand. See equation (2.3.20).

cereal advertising) as well as include store-level promotional variables as covariates. Lastly, I also deflated the advertising expenditures variables which I use by a rough advertising price deflater, in order to control for cyclical changes in advertising costs over the year. None of these changes affected the parameter estimates perceptibly. *Therefore, I conclude that the main results regarding advertising — namely, that $\delta_4 > 0$ and $\alpha_3 < 0$ — are remarkably robust across all specifications, even those (like Model D) which do not employ brand fixed effects.*

The finding that $\alpha_3 < 0$ indicates that, on the margin, advertising *raises* price sensitivities of experienced households. Given that cereals is often classified as a "loss leader" product category,[61] this heightened price sensitivity could affect consumers' choice of which store to shop in. If imperfect competition exists at the retailer level, could manufacturer advertising, then, function as a "coordination device" which alleviates the double marginalization problem in the manufacturer-retailer relationship by promoting retail competition?[62] Lal and Narasimhan [47] develop such a model in which manufacturer advertising intensifies retailer competition.

Further results The main difference between the nested logit (models A, C, D) and PNL (B, E) model results is that the price coefficient α_1 is much smaller in magnitude in the non-PNL models. For example, the point estimate is -1.698 for model C versus -2.018 for model E. This coefficient implies that the PNL model results predict higher price sensitivities for inexperienced households. Elasticity calculation will be presented further below.[63]

Furthermore, in 4 of the 5 specifications (excepting model B), the substitution parameter

[61] A "loss leader" is a product category which retailers discount in order to draw customers into stores.

[62] I have performed some simple regressions utilizing a limited amount of brand-level data on retailer margins for breakfast cereals. The negative correlation between retailer margins and manufacturer advertising suggests to some extent that retailers face more elastic demand with respect to highly-advertised brands.

[63] The PNL models (B and E) were estimated in order to control for measurement error in the past purchases indicator (more specifically, for unobserved purchases). The estimates of the θ parameters are not reported in the tables, but their general qualitative gist is that the estimated unobserved purchase probabilities are rather small, less than 5%. The table here presents these probabilities calculated using the Model E estimates. q_{iht} clearly varies over households and over brands; for instance, households with children have a substantially higher probability of making an unobserved purchase of a kids cereal.

σ is estimated to be positive, indicating lower substitutability between an experienced and an inexperienced brand than between two experienced (or two inexperienced) brands, at the household level. In addition, this σ parameter is larger for households with children, implying that these households are more habit persistent in the sense that they substitute less readily between experienced and inexperienced brands. Referring back to equation (2.2.6), another implication of this finding that $\sigma > 0$ is that experience affects the stochastic portion of utility (i.e., η), so that the included covariates do not completely capture differences between experienced and inexperienced households in their perceptions of a given brand's utility.

Finally, the positive coefficient on PREVPURC in models A and B indicates that households are much more likely to buy cereal in a given week if they have purchased in the previous week. This finding contradicts any "inventory" model of purchase frequency, according to which a household maintains a stock of cereal at home so as not to have to buy cereal each week. However, it is symptomatic of *unobserved heterogeneity* in preferences for cereal across households. Suppose that the sample consists of "normal" and "high" preference households, but this "type" is unobserved, and uncontrolled for by the included household covariates. Then the positive relationship between PREVPURC and current cereal purchase will arise because PREVPURC essentially proxies for households' unobserved types: a "high" type household will (relative to a low type) have a higher propensity to purchase in both the previous and current weeks. Indeed, once unobserved heterogeneity controlled for (in models C and E), the coefficient on PREVPURC becomes very negative (-0.202 in model C, -0.177 in model E).[64]

The sensible negative estimate for the coefficient on PREVPURC certainty favors the specifications (models C, D, and E) which control for unobserved heterogeneity. Furthermore, these two models also have higher log likelihood function values at the converged

Segment:	YGCHIL=0	YGCHIL=1
FAM	1.77%	7.34%
ADULT	1.03%	1.72%
KIDS	1.52%	14.91%

[64] As yet, I have not controlled for the possibility that PREVPURC, like PASTUSE, may be underestimated due to unobserved purchases.

estimates than the other models. This indicates better overall statistical fit of model to data, although comparison of log-likelihood function values should be done with caution since models B and E are estimated via simulation. Nonetheless, in what follows I mostly restrict my attention to the model model E results.

2.6 Implications of the results

2.6.1 Effects of price and advertising on purchase probabilities: an illustrative example

When estimating discrete-choice models, it is often useful to report the findings in terms of probability derivatives $\frac{d\text{Prob}_i}{dx_j}$, where Prob_i is the purchase probability for brand i and x_j is a characteristic of brand j (such as its price or advertising). In the case of the conditional logit model estimated here (i.e., covariates which vary over alternatives, rather than over agents), these probability derivatives are not so straightforward to compute, since the magnitude of these probability derivatives vary depending on which brand the derivative is taken with respect to. To abstract away from brand heterogeneity issues, therefore, I calculate own- and cross-probability elasticities $\frac{d\text{Prob}_i}{dx_j}\frac{x_j}{\text{Prob}_i}$ (with respect to price and advertising) for a hypothetical situation involving choice between 50 *ex ante* identical brands (same price and advertising; they differ only in their η, the stochastic portion of utility). I assume that a given household is experienced with 5 of these brands, and calculate the probability elasticities is response to marginal changes in price and advertising. Furthermore, I distinguish between whether a given brand is experienced or not.

The results for models D and E are reported in figure 2.12. The magnitudes of the results are identical in sign, emphasizing the point made earlier that brand fixed effects (which are missing from model D) affect only the magnitude, but not the qualitative nature of the results. I focus on the model E results here. First, there are large differences in own-price elasticities with respect to brands that are (-2.02) and are not (-5.03) experienced. The much lower price sensitivities towards experienced brands indicates a substantial amount of

Figure 2.12: Own and cross probability elasticities

Calculated for the "illustrative example" described in Section 2.6.1 in the main text.

		Model D results	Model E results
A. With respect to price			
Own elasticities	EXP-EXP[a]	-0.09	-2.02
	INEXP-INEXP	-2.41	-5.03
Cross elasticities	EXP-EXP	0.012	0.225
	INEXP-INEXP	0.034	0.055
	EXP-INEXP	0.004	0.003
	INEXP-EXP	0.002	0.046
B. With respect to advertising			
Own elasticities	EXP-EXP	-0.48	-0.79
	INEXP-INEXP	0.01	0.14
Cross elasticities	EXP-EXP	0.063	0.088
	INEXP-INEXP	-0.000	-0.002
	EXP-INEXP	-0.000	-0.000
	INEXP-EXP	0.012	0.018

Price and advertising set at mean levels: price=\$2.87, advertising=\$2.864 million

[a]This means: elasticity of the purchase probability of an experienced brand with respect to its own price

habit persistence in this market. These differences also exist for the own-advertising elasticities. For inexperienced brands, an increase in advertising raises the purchase probability (the elasticity is 0.14), which is sensible. For experienced brands, however, the negative estimate for α_3 implies that, keeping prices fixed, an increase in advertising *lowers* purchase probabilities (the elasticity is -0.79).[65] In the aggregate (i.e., summing across experienced and inexperienced households) one would expect that the former effect dominates; otherwise these results imply that the average firm is not optimally setting its price and advertising. This is explored in the next section.

Advertising substitutes for experience by making demand for the inexperienced brand more price elastic. Furthermore, the negative estimate for α_3 means that advertising also

[65]The cross effects are less interesting since, in a logit framework, the cross effects are restricted to be opposite in sign from the own effects. Thus, for example, a negative own-price effects translates into positive cross-price effects. The magnitudes of the cross effects, however, reflect the nested logit choice model: note that the "within-nest" cross effects are larger in magnitude than the across-nest cross effects (*e. g.* the cross-price elasticity between two experienced brands is 0.225, and only 0.046 between an inexperienced and an experienced brand).

makes demand for the experienced brand less elastic. These trends are illustrated in figure 2.13, which plots own-price elasticities for an experienced and an inexperienced brand at different values for advertising. There is one caveat here: in making the calculations

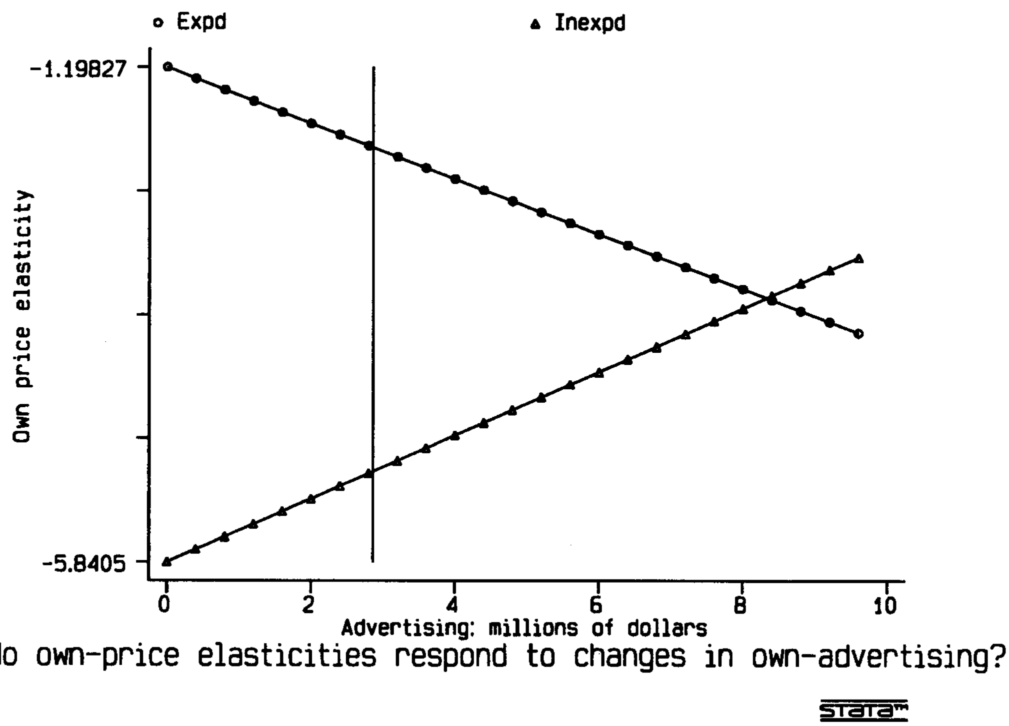


Figure 2.13: Calculated using **Model E** results. **Expd:** own-price elasticities for experienced brand; **Inexpd:** own-price elasticities for inexperienced brand; vertical line at $2.864 million, average quarterly advertising expenditure

used in the graph, it is assumed that prices remain constant as advertising changes. At the household level, this corresponds to varying the household's advertising exposure while keeping the price it faces fixed.[66]

We have seen here at at the household level, a brand's advertising makes an experienced household's probability of purchasing this brand more price elastic, while decreasing the price elasticity of an inexperienced household's purchase probability. At the aggregate level, which effect dominates — Does a brand's advertising increase or decrease that brand's aggregate own-price elasticity? This is examined further below.

2.6.2 Effects of experience and advertising: some calculations

In the previous section I calculated price and advertising elasticities for a hypothetical situation in which all 50 brands are *ex ante* (i.e., before any of them are experienced) identical. In figure 2.14, I compute own-price elasticities using the actual data for the model E results, to show how a given brand's purchase probability, and the own-price elasticities of this purchase probability would change depending on whether a household had recently experienced the brand or not. These elasticities were calculated for, and subsequently averaged over, 220 of the 22146 observations used in the current estimation. These are *short-run elasticities* since, in the calculations, I take a household's purchase history as given. (In the next section, I calculate *long-run elasticities* which account for habit persistence in brand choice behavior.)

Households are much more likely to purchase brands they have recently experienced, and these brands have markedly lower own-price purchase probability elasticities. These differences are quite substantial: for example, experience with a brand lowers a household's

[66]From the graphs, it appears that the own-price elasticities change *linearly* given the increases in advertising. This is a feature of our hypothetical example, and not of the own-price elasticities in general. For our hypothetical example (50 brands, with two nests: either "experienced" or "inexperienced"), the formulas for the own price elasticities are

$$\begin{aligned}\epsilon_{ii} \mid \text{EXP} &= \{\alpha_1 + \alpha_2 * \text{ADV}_i + \delta_2\} * p_i * \left[1 + \sigma * \frac{\exp(V_i)}{\sum_{i' \in \text{EXP}} \exp(V_{i'})} - (1-\sigma) * d_i\right] \\ \epsilon_{ii} \mid \text{INEXP} &= \{\alpha_1 + (\alpha_2 + \delta_4) * \text{ADV}_i\} * p_i * \left[1 + \sigma * \frac{\exp(V_i)}{\sum_{i' \in \text{INEXP}} \exp(V_{i'})} - (1-\sigma) * d_i\right],\end{aligned} \tag{2.6.32}$$

both of which are *nonlinear* functions of ADV_i. However, it appears that the terms inside square brackets ([...]) in the equations above remain largely constant to the changes in advertising, while the terms in curly brackets ({...}) are linear in advertising, with the result that the computed elasticities vary linearly with advertising.

own-price purchase probability elasticity for an average brand from -5.64 to -2.17 (more than a two-fold decrease), while increasing the purchase probability from 0.2% to 9.5% (about a 50-fold increase). Results are similar when looking at average by brand segment, or just for the top 5 brands. These results provide yet more evidence of the extent of habit persistence in household's brand choice decisions.

Figure 2.14: Differences in purchase probabilities and own-price/own-advertising elasticities of purchase probabilities between experienced and inexperienced households

d_i: purchase probability; ϵ_{ii}: own-price elasticity; ϵ^*_{ii}: own-advertising elasticity

	Model E Estimates:	ϵ_{ii} | EXP	d_i | EXP	ϵ_{ii} | INEXP	d_i | INEXP
50 brand avg.		-2.167	0.0950	-5.637	0.0022
Segment:					
Family		-2.403	0.0866	-4.933	0.0035
Adult		-1.983	0.1125	-5.697	0.0021
Kids		-2.296	0.0677	-6.274	0.0008
Top 5 brands:					
Corn Flakes		-2.107	0.1174	-2.631	0.0112
Cheerios		-3.734	0.0961	-4.489	0.0067
Rice Krispies		-3.115	0.1021	-4.594	0.0051
Frstd Flakes		-3.139	0.0793	-3.430	0.0069
KG Rn Bran		-2.371	0.0881	-3.747	0.0051

Elasticities and purchase probabilities calculated for each of the 1010 households in the sample, evaluated at average prices and advertising for each brand. Reported results are averages over all households.

d_i | EXP is average purchase probability for brand i (averaged across the 1010 households). For each household, this purchase probability is calculated assuming that this household is experienced with brand i, as well as the other brands which it bought in the 12 weeks prior to Jan. 1, 1992. Since, for a given household, purchase probabilities for each brand i ($i = 1, \ldots, 50$) are calculated assuming a different set of brands that the household is experienced with, these purchase probabilities will sum to > 1 across brands i.

Analogously for the d_i | INEXP, except purchase probabilities will sum to < 1.

How well does advertising substitute for experience? I quantify this by calculating, for each household h and each brand i, the amount of advertising which would make household h's purchase probability for brand i assuming it is inexperienced with brand i equal to the purchase probability assuming it is experienced but at zero levels of advertising. In other words, for each brand i and each household h, I solve for the level of advertising NEWAD$_i$

which solves:

$$(d_{ih} \mid \mathrm{PASTUSE}_{ih} = 1, \mathrm{ADV}_i = 0) = (d_{ih} \mid \mathrm{PASTUSE}_{ih} = 0, \mathrm{ADV}_i = \mathrm{NEWAD})\,. \quad (2.6.33)$$

Figure 2.15 graphs the actual average observed quarterly advertising expenditures of each brand (the oldad segments of the bars), as well as the median (across all households) value of NEWAD (the mednew segments of the bars) for each brand. Clearly, for most of the brands, the amount of advertising expenditure required to completely duplicate the effects of experience on purchase probabilities exceeds the observed advertising expenditures. This indicates that while advertising does substitute for experience, the effects of experience are still much stronger than those of advertising.

2.6.3 Long-Run Aggregate Brand Purchase Elasticities

Since the brand choice models estimated above are dynamic in the sense that a household's purchase propensities in the future are affected by its choices today, it makes sense to ask what these results predict about the *long-run* effects of changes in price and advertising on household brand choices. Furthermore, given the large differences in responsiveness to price and advertising changes between experienced and inexperienced households, one wonders how these effects play out in the aggregate, across both types of households. Therefore in this section I present estimates of long-run aggregate brand purchase elasticities, calculated using the Model E estimates.

Given that the fundamental phenomena which I model at the micro level are households' brand choices, ignoring quantity decisions altogether, it is not wholly accurate to refer to these as "aggregate demand" elasticities.[67] Nevertheless, in section A.2 in the appendix to this chapter, I compare my long-run aggregate choice choice elasticities to the aggregate demand elasticites presented in recent work by Hausman [37] and Nevo [60] to gauge — albeit in limited fashion — the importance of two factors which distinguish my data (and therefore modeling approach) from theirs: (1) household-level data; and (2) accounting for habit persistence in household purchase decisions.

[67] Goldberg [33] is able to calculate aggregate demand elasticities from her household-level brand choice model estimates since quantity is not an issue in auto demand.

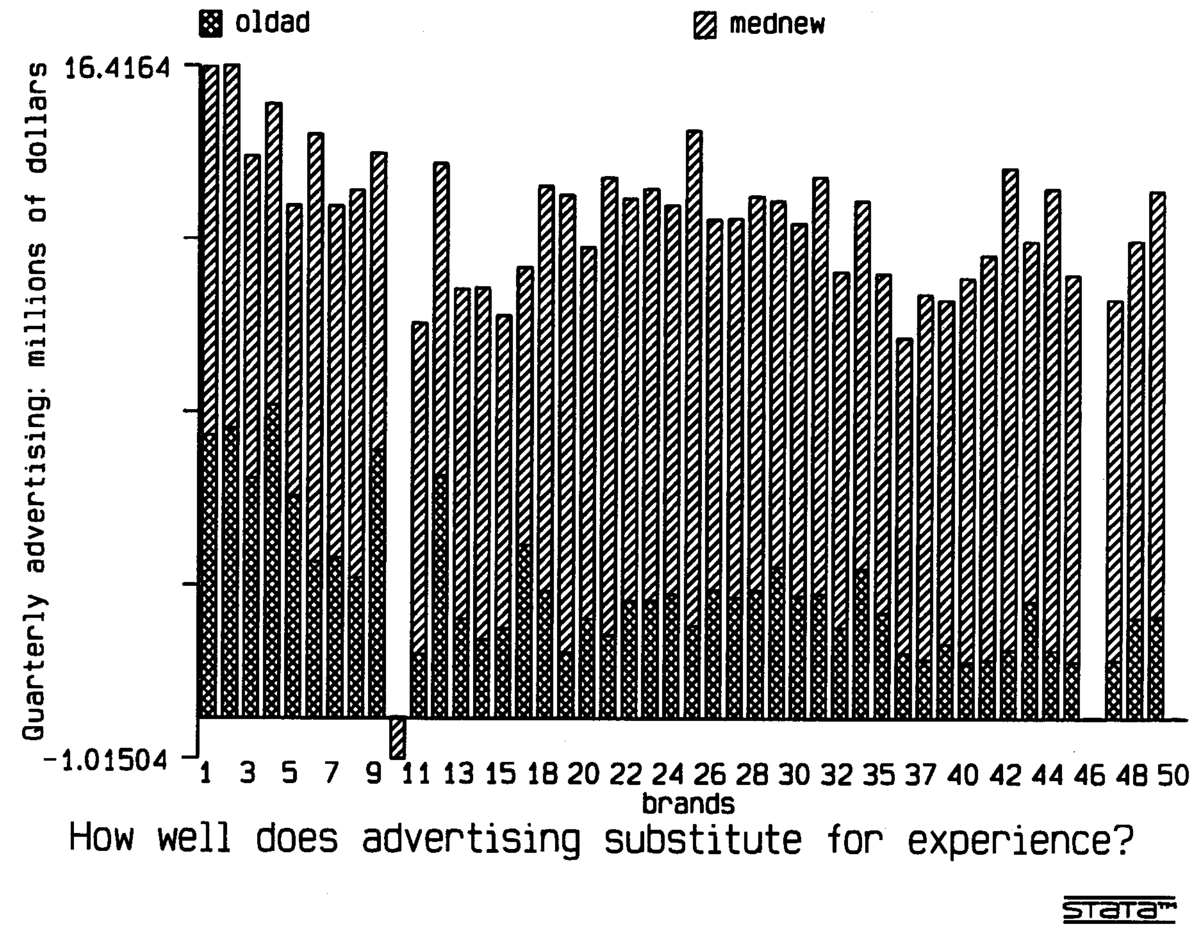


Figure 2.15: How well does advertising substitute for experience? Calculations using Model E results.

oldad: actual average quarterly expenditure, in $millions
mednew: median[a] of NEWAD, quarterly advertising expenditure (in $million) which equates $(d_{ih} \mid \text{PASTUSE}_{ih} = 1, \text{ADV}_i = 0) = (d_{ih} \mid \text{PASTUSE}_{ih} = 0, \text{ADV}_i = \text{NEWAD})$.

Note: From table 2.1, very little was spent on national advertising for brands 10, 46 and 50 during the sample period. From the graph, NEWAD for brands 46 and 50 is also very close to zero, while for brand 10 it is actually *negative*. This latter result is driven in most part by the negative estimate of δ_3 (see main text, section 2.5 for discussion of this finding).

[a]Across all sample households who were not experienced with brand i on 1/1/92.

Full details on the algorithm which I used to calculate long-run aggregate brand choice elasticities are given in section A.1, from the appendix to this chapter; here I just offer a brief summary of the procedure. First I simulate one-year purchase histories for each of the 1010 households in my sample. I aggregate these histories over households in order to estimate long-run (in-sample) aggregate purchases for each brand. Then I derive the long-run *arc* elasticities by calculating how these simulated aggregate purchases would respond to a specified percent increase in price The formula for the price elasticity is:

$$\epsilon_{ij} \equiv \frac{\%\Delta D_i}{\%\Delta p_j} \approx \frac{(D_i(p_j * (1+\Delta_p); p_{k,k\neq j}) - D_i(p_j; p_{k,k\neq j}))/D_i(p_j; p_{k,k\neq j})}{\Delta_p}$$

where D_i and p_j are the long-run aggregate purchases for brand i and the price of brand j, respectively, and Δ_p is a given percentage change in p_j. In the calculations reported here I set Δ_p and Δ_{adv}= to 0.1 (i.e., a 10% price increase). Similar calculations were performed to derive advertising elasticities.

Before presenting the elasticity calculations, I first compare actual observed market shares (of total *purchases*, not quantities) to "long-run" market shares derived from the simulated purchase histories generated using the model E estimates, in order to gauge the overall fit of the model. From the plots in figure 2.16, it appears that the market shares predicted by my model were reasonably similar to observed market shares.[68]

Long-run aggregate price elasticities Figure 2.17 contain the calculated price elasticities using the Model E results. Ten sets of elasticities were simulated, and the reported numbers are the average of these ten estimates. The standard deviations over these 10 sets of estimates are also given. The top part of the figure shows own-price elasticities, both for individual brands as well as averaged over all brands or brand segments. The bottom part of the figure shows the cross-elasticity matrix for the 13 brands in the family segment. The own-price elasticities at the top of figure 2.17 are quite plausible; the average value of

[68]Furthermore, I simulated these purchase histories under several assumptions about "initial conditions": (1) no households experienced with any brand; (2) households experienced only with brands they tried in 12 wks prior to Jan. 1, 1992; and (3) all households experienced with all brands. I also simulated purchase histories for varying lengths of time. Remarkably, predicted aggregate market shares robust to these sensitivity checks. Thus it appears the household-level brand choice model estimated here generates quite realistic aggregate market shares.

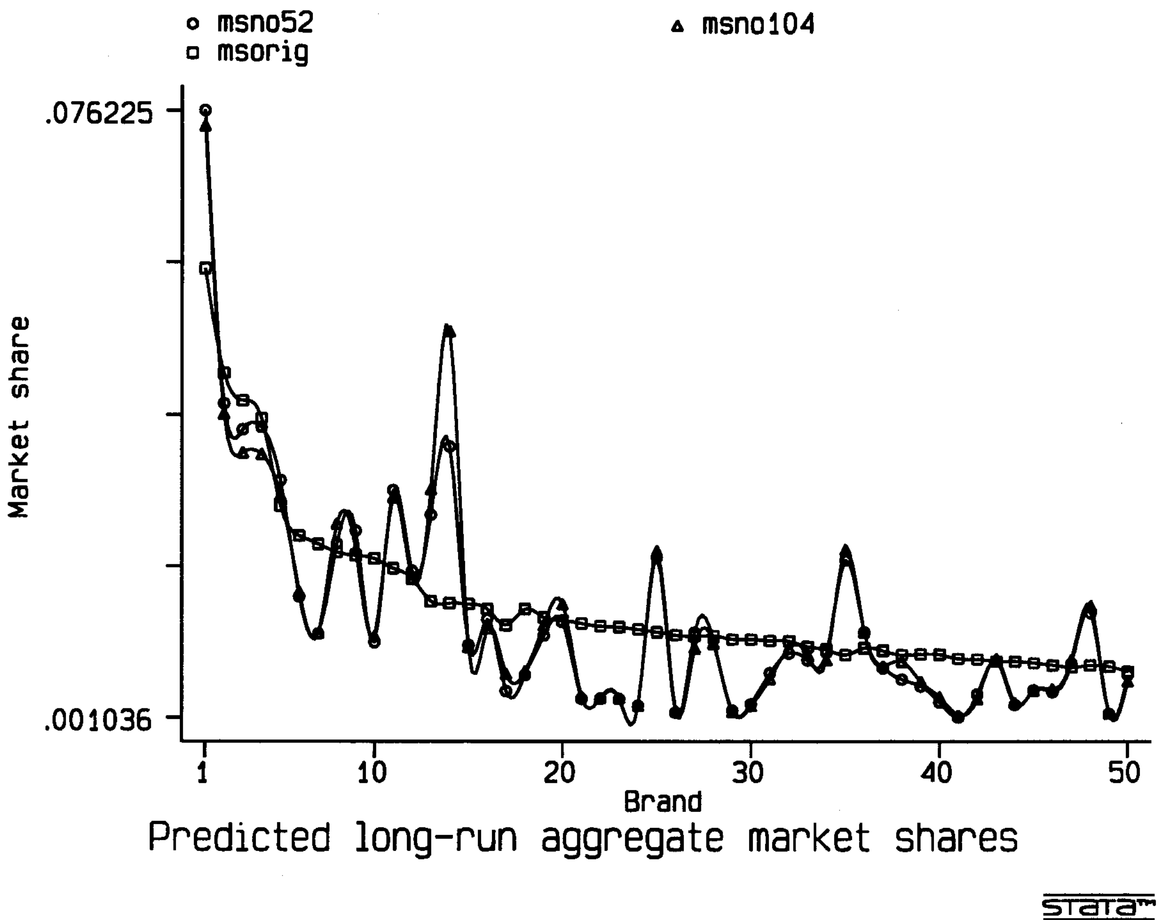


Figure 2.16: Predicted Long-Run Aggregate Market Shares, from Model E Results

msorig: original in-sample market shares
msno52: predicted market shares over 1010 households, calculated simulated 52-week purchase sequences for each household
msno104: predicted market shares over 1010 households, calculated simulated 104-week purchase sequences for each household

Initial conditions: assume households start out experienced with none of the brands

Figure 2.17: Long run price elasticities: Estimated from Model E results

Each household is assumed to make one shopping trip per week, and purchase at most one brand per trip.
In-sample aggregate brand purchases for each brand is the sum (over all sample households) of the estimated number of times each household purchases the brand over 52 weeks.
Arc Elasticities of brand-level aggregate purchases were calculated, for a hypothetical 10% price rise.

Average long-run aggregate own-price elasticities

	Average ϵ_{ii}	St. deviation[a]
Over all brands	-3.832	
FAMILY brands	-3.575	
ADULT brands	-3.848	
KIDS brands	-4.077	
Top 5 brands:		
Corn Flakes	-2.303	0.059
Cheerios	-3.872	0.084
Rice Krispies	-3.763	0.245
Frosted Flakes	-2.824	0.156
KG Raisin Bran	-2.929	0.625

Cross-price elasticity matrix for FAMILY cereals
(first row is average value, second row is standard deviation)

Brand #	1	2	3	4	5	7	15	16	18	31	45	46	48
1: KG	-2.30	0.14	0.13	0.08	0.12	0.03	0.04	0.06	0.03	0.04	0.02	0.03	0.06
CornFl	0.06	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2: GM	0.29	-3.87	0.11	0.16	0.03	-0.05	0.01	0.00	0.04	0.04	0.06	0.01	0.06
Cheerios	0.20	0.08	0.09	0.03	0.04	0.02	0.02	0.02	0.03	0.02	0.03	0.02	0.03
3: KG	0.28	0.23	-3.76	0.06	0.12	0.13	0.19	0.03	0.09	0.06	0.02	0.03	0.05
RiceKr	0.54	0.32	0.24	0.45	0.27	0.15	0.06	0.09	0.04	0.09	0.05	0.05	0.06
4: KG	-0.01	0.11	0.34	-2.82	0.09	-0.09	-0.19	-0.14	-0.13	-0.08	-0.01	-0.03	0.03
FrstdFl	0.35	0.31	0.20	0.16	0.09	0.11	0.06	0.06	0.03	0.07	0.03	0.03	0.04
5: KG	0.53	0.56	0.32	0.45	-2.93	0.05	0.17	0.32	0.26	0.13	0.08	0.06	-0.01
RaisinBn	1.64	0.86	0.92	1.01	0.62	0.23	0.22	0.27	0.16	0.16	0.09	0.05	0.12
7: GM	-1.46	-1.36	-0.77	-0.59	-0.63	-3.98	0.14	0.07	-0.04	-0.27	-0.13	-0.07	-0.11
HN Cheerios	1.51	1.25	1.36	0.62	0.92	0.49	0.16	0.38	0.38	0.44	0.17	0.17	0.42
15: GM	0.77	0.63	0.34	0.31	-0.17	-0.30	-4.00	0.29	0.19	0.18	0.18	0.22	0.20
Wheaties	3.18	1.94	2.68	2.42	2.67	1.74	0.22	1.05	0.44	0.43	0.31	0.35	0.54
16: PT	-0.11	0.01	-0.05	-0.21	0.13	0.01	0.20	-2.75	-0.12	-0.26	-0.11	-0.20	-0.60
RaisinBn	2.24	1.24	2.22	1.63	1.70	0.28	0.26	0.49	0.34	0.22	0.34	0.22	0.61
18: KG	-0.02	0.11	-0.08	0.33	0.74	0.17	0.10	0.36	-4.57	-0.20	0.43	0.39	0.05
CornPops	2.90	5.93	2.81	5.84	3.77	2.26	3.97	3.19	0.46	1.13	0.78	0.63	2.17
31: GM	0.28	-0.01	-0.09	-0.27	-0.09	0.28	-0.06	-0.42	-0.29	-4.08	0.33	0.41	0.72
AC Cheerios	7.82	2.38	2.37	2.37	2.66	2.44	3.81	2.81	2.36	0.91	0.55	0.38	1.43
45: GM	0.14	0.65	0.44	0.42	0.00	-0.05	0.09	0.34	-0.23	-0.04	-3.63	-0.42	0.77
Clusters	16.05	26.34	8.60	9.68	6.60	5.62	4.63	4.82	3.52	3.04	1.25	2.10	9.61
46: KG	0.34	0.17	-0.06	0.57	0.40	0.19	0.40	0.25	0.03	-0.30	-0.49	-3.77	1.00
MiniBuns	7.02	10.97	12.02	11.91	12.66	3.66	2.56	4.37	2.27	1.37	0.97	0.84	5.59
48: GM	0.23	0.81	1.12	1.03	0.81	1.15	0.95	0.99	0.80	0.92	0.62	0.84	-4.00
MG Cheerios	5.50	6.20	2.02	3.03	2.45	1.24	0.73	0.97	1.59	0.72	0.47	0.47	0.66

[a]Average and standard deviation calculated over 10 simulated purchase histories

-3.832 corresponds, roughly speaking, to a price-cost margin between 25% and 30%, using the monopolist's inverse-elasticity pricing rule and ignoring oligopolistic competition and multiproduct pricing.

Previously, I demonstrated that at the household level, advertising makes an experienced household's brand purchase probability more elastic, while decreasing the elasticity of an inexperienced household. Figure 2.18 shows that, aggregating across both inexperienced and experienced households, the latter effect dominates for four out of the fifty brands, using the estimates from model E. In other words, advertising *reduces* the own-price elasticities of aggregate purchases. This is the same effect of advertising as in the standard "reduced-form" models of persuasive advertising.[69]

The main caveat in these calculations is, as before, that prices are assumed to remain constant while advertising expenditure varies. This assumption is justified to the extent that prices are set by retailers who needn't respond to changes in national advertising levels. Alternatively, one can think of the trends plotted in figure 2.18 as descriptive of the effects of advertising in a world in which prices are regulated — according to these results, the net effect of advertising in the aggregate is to reduce the own-price elasticities of brand purchases.[70]

While in the static logit model the cross-price elasticities are restricted to be positive (as long as the price coefficient is negative), here we see that many of the long-run cross-price elasticities are negative, indicating long-run *complementarities* among brands. However, it is very difficult to draw any robust conclusions from the cross-price elasticities. In general, these numbers are very small in magnitude, and furthermore not at all precisely (as measured by the standard deviations) estimated, except for the largest brands (those in the top rows of the matrix).

[69]In a linear demand model where advertising enters positively, a rise in advertising will decrease the own-price elasticity of demand, holding price fixed.

[70]On the other hand, the empirical analysis of cigarette advertising performed by Roberts and Samuelson [64] likewise assume away price changes in simulating the effects of changes in advertising. It is beyond the focus of this chapter to fully develop a supply-side model describing firms' joint choices of price and advertising. Such a model would prove an interesting extension, however.

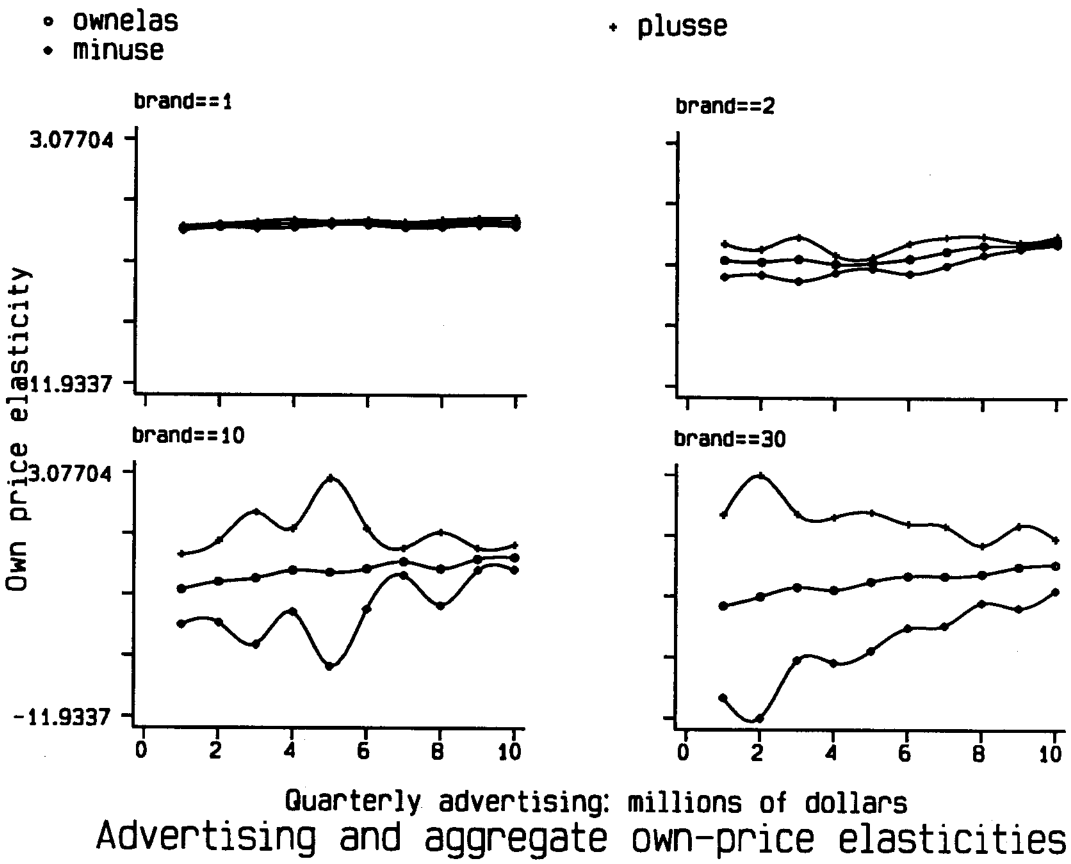


Figure 2.18: Calculated using **Model E** results. **ownelas:** own-price elasticities; **plusse:** +2 standard deviations; **minuse:** -2 standard deviations

brand=1: KG Corn Flakes (average quarterly AD$=7.11); **brand=2:** GM Cheerios (average quarterly AD$=7.29); **brand=10:** NB SpoonSize Shredded Wheat (average quarterly AD$=0.03); **brand=30:** GM Lucky Charms (average quarterly AD$=3.08)

Caveat: Throughout, assume price remains constant while advertising level is varied

Figure 2.19: Long run advertising elasticities: Estimated from Model E results

Each household is assumed to make one shopping trip per week, and purchase at most one brand per trip.
In-sample aggregate brand purchases for each brand is the sum (over all sample households) of the estimated number of times each household purchases the brand over 52 weeks.
Arc Elasticities of brand-level aggregate purchases were calculated, for a hypothetical 10% advertising rise.

Average long-run aggregate own-advertising elasticities

.	Average ϵ_{ii}	St. deviation[a]
Over all brands	0.044	
FAMILY brands	-0.006	
ADULT brands	-0.075	
KIDS brands	0.348	
Top 5 brands:		
Corn Flakes	-0.365	0.106
Cheerios	-0.176	0.188
Rice Krispies	-0.297	0.189
Frosted Flakes	0.070	0.531
KG Raisin Bran	-0.306	0.077

Cross-price elasticity matrix for FAMILY cereals
(first row is average value, second row is standard deviation)

Brand #	1	2	3	4	5	7	15	16	18	31	45	46	48
1: KG	-0.36	0.05	0.11	0.16	0.04	0.02	0.01	0.01	0.00	0.02	0.00	0.00	-0.01
CornFl	0.11	0.11	0.10	0.09	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2: GM	-0.04	-0.18	-0.17	-0.27	0.09	-0.07	0.02	0.04	0.01	0.00	0.02	0.00	0.06
Cheerios	0.40	0.19	0.30	0.23	0.09	0.02	0.01	0.01	0.03	0.01	0.00	0.00	0.01
3: KG	0.24	-0.07	-0.30	-0.15	-0.19	-0.15	-0.05	-0.09	-0.04	-0.04	0.02	0.00	0.03
RiceKr	0.61	0.78	0.19	0.26	0.12	0.12	0.01	0.02	0.04	0.01	0.00	0.00	0.02
4: KG	0.20	0.42	0.37	0.07	0.13	0.21	0.03	0.07	-0.01	-0.03	-0.02	0.00	-0.01
FrstdFl	0.29	1.27	0.39	0.53	0.20	0.12	0.02	0.16	0.02	0.01	0.01	0.00	0.04
5: KG	-0.10	-0.17	-0.30	0.03	-0.31	-0.18	-0.01	-0.14	-0.03	0.02	-0.04	0.00	-0.05
RaisinBn	0.42	0.48	0.26	0.29	0.08	0.06	0.01	0.17	0.05	0.01	0.01	0.00	0.02
7: GM	0.30	0.11	-0.06	0.54	0.13	0.57	-0.03	0.04	-0.09	0.17	0.04	0.00	0.03
HN Cheerios	1.20	1.36	2.13	1.04	1.00	2.20	0.07	0.04	0.28	0.05	0.04	0.00	0.14
15: GM	0.75	0.53	-0.05	0.44	-0.08	-0.59	-0.44	-0.30	-0.46	-0.44	0.05	0.00	-0.23
Wheaties	5.00	4.17	2.71	4.51	1.94	1.79	0.91	1.05	0.47	0.72	0.06	0.00	0.28
16: PT	0.21	-0.21	-0.07	-0.46	0.14	0.13	-0.10	-0.02	-0.17	-0.06	0.01	0.00	-0.06
RaisinBn	3.26	4.24	2.21	1.88	0.98	0.22	0.05	0.14	0.11	0.08	0.18	0.00	0.10
18: KG	0.43	0.59	0.75	0.82	1.24	0.47	0.45	0.61	0.57	0.02	0.06	0.00	-0.06
CornPops	1.42	4.59	6.65	6.56	3.87	2.02	0.51	1.58	0.82	0.79	0.05	0.00	0.28
31: GM	0.46	1.36	0.21	0.89	0.01	0.58	0.18	0.71	0.68	0.14	0.25	0.00	-0.07
AC Cheerios	8.23	8.35	3.14	3.00	6.73	4.04	1.57	3.47	4.15	5.20	0.90	0.00	0.14
45: GM	-1.37	3.35	1.17	3.31	0.69	0.58	0.29	0.85	-0.11	-0.14	0.00	0.00	0.18
Clusters	6.84	18.22	16.61	50.60	4.83	3.59	1.88	1.02	1.54	1.26	0.52	0.00	1.02
46: KG	-0.47	-0.58	-0.05	-0.24	0.55	-0.25	-0.16	-0.05	-0.07	0.02	-0.05	0.00	-0.03
MiniBuns	8.80	17.74	9.15	18.52	7.24	2.04	0.28	0.74	0.42	0.58	0.08	0.00	0.68
48: GM	0.73	0.11	0.01	0.41	0.60	0.32	-0.07	0.07	0.08	0.31	0.00	0.00	0.18
MG Cheerios	1.84	4.43	1.72	17.42	1.31	0.07	0.16	0.23	0.28	0.25	0.04	0.00	0.38

[a]Average and standard deviation calculated over 10 simulated purchase histories

Long run aggregate advertising elasticities Long run advertising elasticities for the Model E results are presented in figure 2.19. On the whole they are not very precisely estimated (as evidenced by the large standard deviations). While the average own-advertising elasticity across the top fifty brands is small and positive (0.044), as we would expect, the average own-advertising elasticities are actually *negative* for family and adult brands of cereal, as well as for 4 of the top 5 brands of cereal. The subsequent implication is that for many brands, advertising is occurring at excessive levels, which *decrease* aggregate purchases on the margin.

Clearly, this result is driven by the negative estimate for α_3, which is robust across various specifications. However, in performing these simulations I assume that cereal prices remain constant over time. To the extent that the negative estimate on α_3 is due to the fact that households purchase their experienced brands during weeks when those brands are relatively "cheaper",[71] by not allowing relative prices among brands to change over time I may be underestimating households' purchase probabilities for brands they have experienced.

2.7 Conclusions

The major finding of this chapter with regard to advertising — that it increases price sensitivities for experienced households, but "mimics" experience by decrease price sensitivities for inexperienced households — implies that advertising is potentially much more useful for *new* rather than existing brands. That is to say, *assuming price remains fixed*, a marginal increase in advertising expenditures yields more revenue for the entrant than an incumbent, since more households are already familiar (read: experienced) with the incumbent's product. This corroborates Shapiro's [66] intuition that advertising cannot be a barrier to entry if it substitute[s] for, or mimics the effects of, experience when consumers' preferences are characterized by habit persistence.[72]

[71] Indeed, I have run some exploratory reduced-form regressions which indicate that this may be true to a certain extent.

[72] Advertising could as well be found to "complement" experience. Advertising of this sort can be thought of as confirming or "framing" experience, validating to some extent the experiences that a consumers has

Again keeping prices fixed, another implication is that fewer brands may exist in the cereals market in the absence of advertising. Unfortunately, I cannot make the stronger statement that the market would be *more* concentrated (i.e., fewer firms would exist) without advertising, since this market is dominated by several large multiproduct[73] firms. Since this analysis was done at the brand level, I can only conclude that my results indicate that a producer would find it more profitable to advertise a new rather than an old brand; alternatively, that advertising helps new brands gain market share by creating "synthetic experience". It is difficult to generalize about whether this new brand would tend to be introduced by an entrant or an incumbent without making some sweeping assumptions about production costs. Economies of (brand) scope — whether in production or in advertising — would tend to favor the incumbent.

had with a product (see Deighton, et. al. [29] and sources therein for more on these hypotheses). The effect of such advertising actually increases with experience, thus reinforcing habit persistence and therefore experience-bred product differentiation.

[73]Actually, *multimarket* firms. The 3 largest cereal producers (Kelloggs, General Mills, Phillip Morris (Post)) also compete with each other across other food markets.

Chapter 3

Informational Barriers to Entry in the Anti-Ulcer Drug Market: the Case of *Omeprazole*

Much of the previous chapter was motivated by Shapiro's argument that advertising can actually promote competition in a market where brand loyalty — built up through first-hand experience with a product — is important. Although we did not emphasis this aspect of his argument previously (since it was not relevant there), Shapiro clearly attributes the importance of first-hand experience in a differentiated product market to lack of information among consumers; he states ([66], pg. 7)

> ...the fundamental source of the entry barrier is an information one: consumers have better information about established brands than about new ones [...] information is the basic barrier to be overcome by a new product...

A quite sizeable theoretical literature (see Schmalensee [65] and Bagwell [7] for two examples) have developed this idea, and drawn out the welfare implications of these "informational barriers to entry".

In this chapter we empirically verify the existence of and quantify the extent of informational barriers to entry into the anti-ulcer drug market, the largest therapeutic drug market worldwide. We study the diffusion process of a new anti-ulcer molecule (*omeprazole*) over time by using a panel dataset showing all the prescriptions written by a sample of doctors in Rome to all their patients during a three year period (1990-1992). *Omeprazole*, patented by Astra, and marketed worldwide under the brand name LOSEC (but in the U.S. as PRILOSEC), became the world's best selling drug in 1996, grossing more than 2bn$, and gaining about half of the worldwide anti-ulcer market.[1] However, this market ascendance was gradual; during the sample period, drugs based on *omeprazole* managed to garner only 20% of the Italian market, and even in mid-1995, it still clearly lagged behind the H_2-antagonist molecules in terms of market share, measured as percentage of total anti-ulcer drug revenues (see figure 3.1).[2]

To what extent is this gradual growth in market share attributable to doctors' reluctance to prescribe the molecule, due to lack of information about its quality? Furthermore, how important is first-hand experience with *omeprazole* (i.e., actual prescriptions of the molecules to patients) in informing doctors about its quality, relative to other possible sources of information, such as marketing activities undertaken by the manufacturers of the new *omeprazole*-based drugs? To answer these questions, we specify a learning model in which doctors, initially uncertain about the quality differential between *omeprazole* and the incumbent molecules, update their beliefs about this differential after observing, through first-hand experience, noisy signals of it from patients to whom they have prescribed the molecule. The component of *omeprazole*'s rise in market share which we can explain with this learning model provides evidence of the existence and extent of informational barriers to entry in this market: the time required for the doctors to accumulate the requisite first-hand information can potentially reduce the discounted future stream of profits by enough

[1] This information comes from the websites of Astra-Merck and Glaxo.

[2] Each molecule in the Italian market is usually marketed by more than one firm. Usually one of the firms is the patent-holder and the others are licensees. *Omeprazole* was introduced and jointly marketed by three firms (*Bracco*, *Malesci* and *Schering Plough*), none of which was the patent-holder. Curiously enough, *Astra*, the patent-holder, only entered the Italian market in May 1993 (after the end of our sample period), because it did not have a marketing division in Italy at the time *omeprazole* was patented. Therefore *Astra* decided to license the product before directly entering the market. There are interesting competitive issues between the patent-holder and the licensee to be analyzed, but we assume them away and focus on competition *across* molecules.

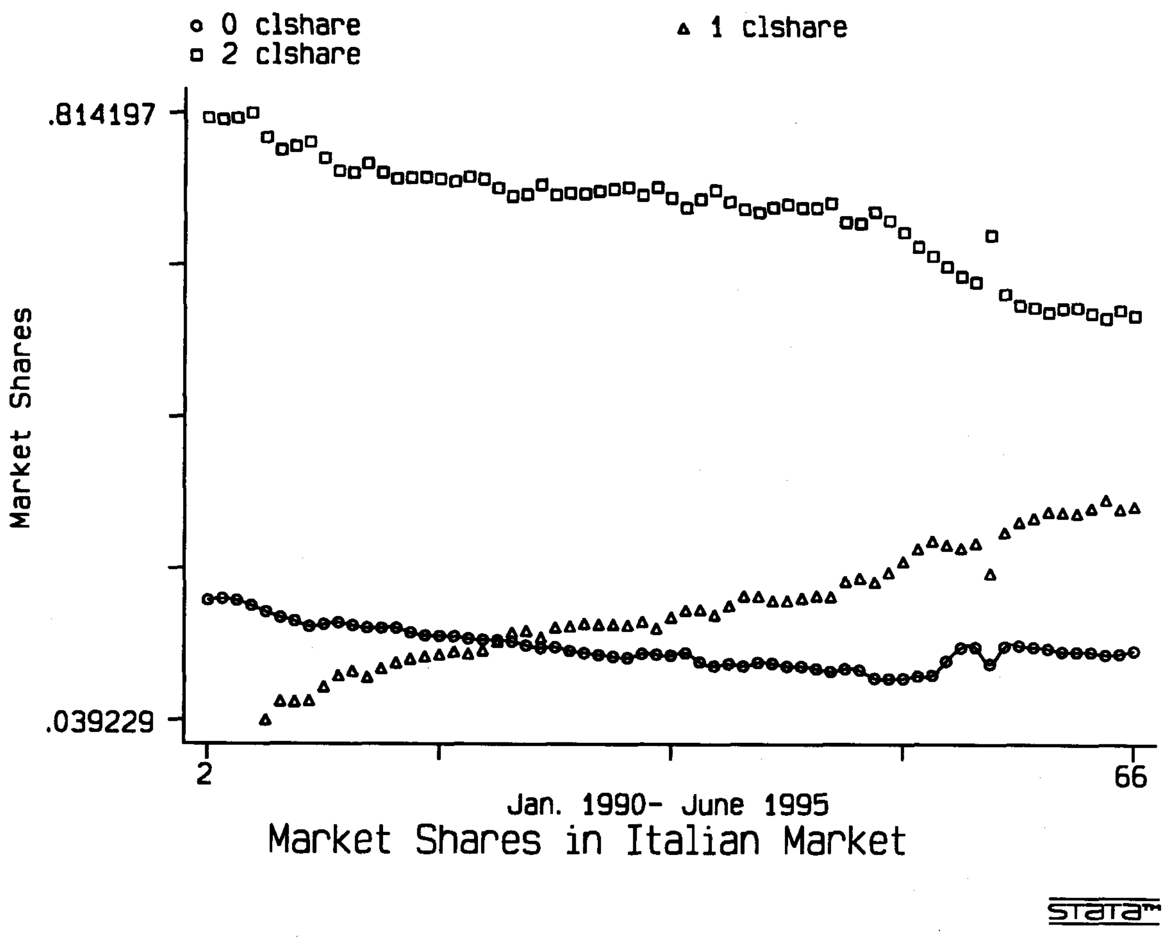


Figure 3.1: Market shares in the Italian anti-ulcer drug market
Market shares measured as percentage of total anti-ulcer drug revenues, in a given month.

to deter firms from introducing new drugs which would have been profitable in a perfect information world.

We find that doctors' initial pessimism about *omeprazole*'s quality is largely responsible for the gradual growth in prescriptions of it over the sample period. This finding is not too surprising, given that the learning model rationalizes the positive serial correlation in doctor-level prescriptions of *omeprazole* observed in the data as initial pessimism. However, our striking finding is that this uncertainty is resolved by first-hand experience (actual prescriptions) rather than through marketing activities and other potential aggregate sources of information (such as journal publications and word-of-mouth among doctors): this despite including a complete set of dummies for all the months in our sample in order to capture these non-experience aggregate effects.

In other words, we find that the gradual rise in market share in our data is driven by *doctor-level*, rather than aggregate (across all doctors) serial correlation in prescriptions of *omeprazole* . This is strong evidence in favor of the learning model in which, similarly, movements in aggregate market share are driven by doctor-level learning rather than by aggregate effects. Our results are very striking here: the predicted prescription probability remains very low (around 2%) and virtually unchanged over the sample period for doctors without first-hand experience, whereas for a doctor who prescribes *omeprazole* an "average" proportion of the time (relative to the sample), it rises from about 2% to over 20% over the sample period: thus learning through first-hand experience accounts for almost *all* the observed rise in market share over the sample period!

3.1 Background

The doctor/patient relationship is fraught with uncertainty:[3] doctors have incomplete information on the medical condition of a patient, and which treatment is best for the patient. How do doctors learn about the quality of alternative treatments? Oftentimes, only through

[3]For a general discussion of these issues, see Phelps [63].

direct experience – actual prescriptions of the new molecule. This first-hand experience accumulates slowly, however, and is confounded by heterogeneity across diagnoses: what works for diagnosis X may not work as well for diagnosis Y. Otherwise, doctors can get some secondhand information (such as advertising and sales pitches by the manufacturers of medical treatments, articles in medical journals, or recommendations from other doctors). This uncertainty at the individual physician level can have significant implications for competition and market structure in pharmaceutical markets. If new, innovative medical treatments face informational hurdles in gaining acceptance among doctors, then existing treatments enjoy a "pioneer advantage".[4] The disadvantages faced by entrants' products due to doctors' lack of information about them — we call these informational barriers to entry — are the focus of our analysis.

Several studies have documented the existence and, more importantly, the nature of barriers to entry into pharmaceutical markets. Early research by Bond and Lean [17] investigated the phenomenon of pioneer advantage, and reasons under which it persists or erodes in two therapeutic drug markets: diuretics and antiaginals. Not surprisingly, they find substantial pioneer advantage. More interestingly, however, they also find that late entrants offering products containing some therapeutic novelty managed to gain large market shares when backed by heavy advertising campaigns. On the other hand, late entrants offering products of similar quality (i.e., "me-too" brands) failed to gain large market shares, even after expensive advertising campaigns.

Like Bond and Lean, Berndt, Bui, Reiley and Urban ([12], hereafter BBRU) measure the importance of *marketing* in overcoming pioneer advantage in the anti-ulcer drug market. Like the work in the previous chapter, they measure the effect of advertising in raising demand for newer, less familiar, brands, but they look for market-level effects, whereas my analysis in the last chapter focused on household-level effects.

Theirs in an econometric analysis of the rise to market dominance of the Glaxo's late

[4]which can be quantified (or "pecuniarized") in terms of the extra (relative to the incumbents) promotional costs which entrants must incur in order to capture a given market share.

entrant ZANTAC (*ranitidine*), especially vis-a-vis the pioneer and early market leader TAGAMET (*cimetidene*, marketed by Smith-Kline), in the anti-ulcer drug market. Their findings show quite clearly that product advances (such as those embodied in ZANTAC) don't necessarily translate into success in therapeutic drug markets (in terms of market share) without tremendous marketing muscle. As striking as this result is, however, they never explain the causes of pioneer advantage. Given that there is a sizeable theoretical economics literature (cited earlier above) on informational barriers to entry, the availability of doctor-level prescription histories allows us a unique opportunity to quantify the extent of this informational pioneer advantage. Furthermore, we are able to compare the effectiveness of first-hand experience versus promotional efforts in convincing doctors to prescribe the entrant drug, issues which eluded BBRU due to data constraints.

The anti-ulcer drug market: some background Being the largest therapeutic drug market worldwide, the anti-ulcer drug market is segmented across different diagnoses (submarkets), with preferred treatment differing across segments depending on the severity of the diagnosis. These segments are summarized in figure 3.2 and discussed below.

Figure 3.2: **Diagnoses and Treatments**

Diagnosis	**Treatment**[a]	**Preferred drugs**[b]
1—Minor heartburn	Drugs or no Prescription	"Low-Tech" drugs
2—Pathological hypersecretory conditions	Anti-ulcer drugs	*omeprazole*
3—Attack therapy for GERD or Peptic Ulcer	Anti-ulcer drugs	*omeprazole*
4—Maintenance therapy for GERD or Peptic Ulcer	Anti-ulcer drugs	H_2-antagonists

[a] *Physicians' Desk Reference* [56].
[b] *Physicians' Desk Reference* [56].

The two most common diagnoses requiring treatment using anti-ulcer drugs are peptic ulcers and Gastroesophagal Reflux Disease (GERD). **Peptic ulcers** strike more than 10% of all Americans at one time or another during their lives. An ulcer is a focal area of the stomach (*gastric*) or duodenum (*duodenal*) that has been destroyed by digestive juices and stomach acid. Gastroesophageal Reflux Disease (GERD) refers to the backward flow of acid from the stomach up into the esophagus. During the 1990-92 period covered in our sample, traditional treatments for these two diagnoses consisted in attacking the ulcer either with H_2- receptor antagonists (the sample-period market leader *ranitidine*, as well as *cimetidine*, *nizatidine*, *famotidine* and *roxatidine* belong to this class), or protonic pump inhibitors (*omeprazole* was the first approved molecule in the latter class). After healing occurred, the patient would either stop the treatment hoping that no recurrence would occur, or keep receiving an anti-ulcer drug for a "maintenance" therapy.

The third main diagnosis is a pathological hypersecretory condition: different syndromes are defined under this label (e.g., Zellinger-Ellison).[5] Finally, on occasion anti-ulcer drugs are also prescribed for minor heartburns (which are, however, often symptomatic of more serious gastrointestinal disorders).

Drugs based on the molecule *omeprazole* entered the market in mid-1990. *Omeprazole* inhibits the gastric acid ("protonic") pump, thereby directly blocking acid production, irrespective of the acid-producing stimulus. The older H_2- receptor antagonist molecules (such as *ranitidine*, the active ingredient in then-market leader ZANTAC, as well as *cimetidine* the active ingredient in the pioneer H_2-antagonist based drug TAGAMET), which only inhibit the acid-producing actions of histamine-2 (H_2 for short) at its receptor cells, but largely ineffective in suppressing acid production arising from other stimuli.

Omeprazole was found in clinical trials to be more effective than H_2- receptor antagonists in "attacking" serious ulcers and severe esophagitis (a GERD-related condition). On the other hand, other clinical studies on animals found that *omeprazole* produced a dose-related increase in gastric carcinoid tumors. Astra-Merck [6] recommends that a standard treatment involving Prilosec last no longer than 8 weeks, making it inappropriate for maintenance

[5] Moreover a cure has not been found yet, so that right now only the symptoms are treated.

therapies. Continuous use of *omeprazole* was advised only for the pathological conditions (e.g., Zellinger-Ellison) where H_2- receptor antagonists were ineffective (Physicians' Desk Reference [56]). Furthermore, the molecule is too strong for heartburns.

Beginning around 1993 (just after the sample period of our data), researchers found that many duodenal ulcers[6] are due to the presence of a bacterium–*Helicobacter Pylori*. The use of antibiotics to fight the *H.Pylori* infection is a major scientific advance; recent studies show that antibiotics can permanently cure 80-90 percent of peptic ulcers (Graham [34]). Despite these advances in curing ulcers, the anti-ulcer drug market has not shrunk appreciably in recent years; in fact, antibiotic treatments for ulcers often involve use of an anti-ulcer drug in combination.[7] Furthermore, antibiotics are not effective against maladies not caused by *H.Pylori*; these include not only gastric ulcers, but also GERD and pathological hypersecretory conditions.

According to Astra-Merck's 1996 annual report [6], 50% of new Prilosec prescriptions (in the U.S.) were for GERD, 16% for ulcers, and 34% for other conditions.

3.2 The Learning Model

There has been a long empirical literature in marketing examining diffusion patterns for new products (see Bass et. al. [9] for a review). Our analysis links these models – mostly formulated at the macro-level – to micro-level demand data. Motivated by the literature on informational barriers to entry, we focus on the accumulation of first-hand experience at the micro- (individual doctor) level to explain the macro-level diffusion pattern.

3.2.1 Doctor's Problem

Doctor i's decision problem involves two steps: first a diagnosis ($j \in J$) is formulated, then a treatment (molecule) is chosen. We model the second step, since the formulation of a diagnosis does not have an economic dimension. We distinguish between different

[6]90%, according to Astra-Merck's (patent-holders of *omeprazole*) 1996 annual report [6]

[7]As Astra-Merck proudly note in [6], Prilosec in combination with the antibiotic Biaxin (clarithromycin) was the first FDA-approved treatment for *H.Pylori* infections.

diagnoses in our model because, as described above, anti-ulcer drugs differ in effectiveness and suitability across diagnoses.

The process whereby the doctor learns about the quality of the new molecule is specified as restrictions on doctor i's choice problem: roughly speaking, at time t she always prescribes the "optimal" molecule depending on the information available. More precisely, at time t, doctor i decides whether or not to prescribe *omeprazole* to a patient k with diagnosis $j \in J$. In making this choice, doctor i chooses the option with the higher utility. Throughout, agency problems between the doctor and the patient are abstracted away, so that we assume the doctor maximizes the *patient's* utility from the prescription. The reputation effects concomitant with the long-term nature of many patient–doctor relationships in Italy (the National Health Service requires each enrollee to list a general practitioner) minimize the divergence between doctors' and patients' objective function which form the basis of agency problems.

We assume that doctor i distinguishes between *two* alternatives: the new molecule, *omeprazole* (alternative 1), and *any* of the other molecules (alternative 0). The utilities from each alternative (for patient k with diagnosis j during period t) are:

$$U^i_{1jkt} = \alpha^*_1 x^i + \beta p_{1t} + \gamma \mathrm{ADV}_{1t} + \xi^*_{1t} + \delta^*_{1j} + \epsilon^i_{1jkt} \quad \text{if new molecule} \tag{3.2.1}$$

$$U^i_{0jkt} = \alpha^*_0 x^i + \beta p_{0t} + \gamma \mathrm{ADV}_{0t} + \xi^*_{0t} + \delta^*_{0j} + \epsilon^i_{0jkt} \quad \text{if any other molecule} \tag{3.2.2}$$

where

- the vector x^i contains doctors' characteristics
- p_{1t} and p_{0t} are, respectively, the price of *omeprazole* and a weighted average of the prices of the incumbent drugs weighted by their market shares at time t.
- ADV_{1t} and ADV_{0t} are, respectively, advertising for *omeprazole* and the *sum* of advertising by the incumbent brands, during period t.
- ξ_{1t} and ξ_{0t} parameterize period t factors which affected the attractiveness of, respectively, *omeprazole* and the incumbent drugs. These are the same over all doctors, patients, and diagnoses. In particular, ξ_{1t} proxies for aspects of the learning process which we don't explicitly model, such as word of mouth and journal articles.

- δ_{1j} and δ_{0j} parameterize the "unobserved quality" of *omeprazole* and the incumbent drugs, respectively. These are unobserved by the econometrician. Doctors, however, are presumed to know δ_{0j}, and have imperfect information about δ_{1j}. How doctors learn about δ_{1j} is described further below.
- ϵ^i_{1jkt} and ϵ^i_{0jkt} are *i.i.d.* (over doctors, patients, diagnoses, and time periods) shocks associated with, respectively, *omeprazole* and the incumbent drugs. They are observed by the doctors, but not by the econometrician.

The choice rule for the doctor is:

$$\begin{gathered}\text{prescribe } \textit{omeprazole} \text{ iff} \\ E_t(U^i_{1kjt}) > U^i_{0kjt},\end{gathered} \tag{3.2.3}$$

which involves an expected value because δ_{1j} is a random variable for doctor i, for all diagnoses j. Since U^i_{1kjt} is linear in δ_{1j}[8], the choice rule (3.2.3) can be rewritten as: that the doctors compare are:

$$\begin{aligned}&\text{prescribe } \textit{omeprazole} \iff \\ &\alpha_1^* x^i + \beta p_{1t} + \gamma \text{ADV}_{1t} + \xi^*_{1t} + E_t\delta^*_{1j} + \epsilon^i_{1jkt} > \alpha_0^* x^i + \beta p_{0t} + \gamma \text{ADV}_{0t} + \xi^*_{0t} + \delta^*_{0j} + \epsilon^i_{0jkt} \longleftrightarrow \\ &\left(\epsilon^i_{1jkt} - \epsilon^i_{0ijkt}\right) > \\ &\qquad -\left((\alpha_1^* - \alpha_0^*)x^i + \beta(p_{1t} - p_{0t}) + \gamma(\text{ADV}_{1t} - \text{ADV}_{0t}) + (\xi^*_{1t} - \xi^*_{0t}) + (E_t\delta^*_{1j} - \delta^*_{0j})\right).\end{aligned} \tag{3.2.4}$$

Given our assumption that ϵ^i_{1jkt} and ϵ^i_{0ijkt} are independent, and each distributed extreme value, the probability that doctor i prescribes *omeprazole* takes the binary logit form:

$$\begin{aligned}&Prob(\text{prescribe } \textit{omeprazole}\ ) = \\ &\frac{\exp\left((\alpha_1^* - \alpha_0^*)x^i + \beta(p_{1t} - p_{0t}) + \gamma(\text{ADV}_{1t} - \text{ADV}_{0t}) + (\xi^*_{1t} - \xi^*_{0t}) + (E_t\delta^*_{1j} - \delta^*_{0j})\right)}{1 + \exp\left((\alpha_1^* - \alpha_0^*)x^i + \beta(p_{1t} - p_{0t}) + \gamma(\text{ADV}_{1t} - \text{ADV}_{0t}) + (\xi^*_{1t} - \xi^*_{0t}) + (E_t\delta^*_{1j} - \delta^*_{0j})\right)} \equiv \\ &\frac{\exp\left(\alpha x^i + \beta \Delta p_t + \gamma \Delta \text{ADV}_t + \xi_t + E_t\delta_j\right)}{1 + \exp\left(\alpha x^i + \beta \Delta p_t + \gamma \Delta \text{ADV}_t + \xi_t + E_t\delta_j\right)}\end{aligned} \tag{3.2.5}$$

where we have substituted $\alpha \equiv \alpha_1^* - \alpha_0^*$, $\xi_t \equiv \xi^*_{1t} - \xi^*_{0t}$, and $E_t\delta_j \equiv E_t\delta^*_{1t} - \delta_{0t}$. We estimate α, γ, and the ξ_t's. The central question in this chapter is whether *omeprazole*'s gradual

[8] We do not assume risk aversion in the utility function, because it is not possible to separately identify the risk aversion parameter and the mean of the initial distribution of the "true" quality.

increase in market share over time is better explained by doctor-specific learning through prescriptions (which enters the model through changes in $E_t\delta_j$) or through secondhand sources of information (captured by ADV and the ξ_t's). The learning process explained in the next section.

3.2.2 Learning Process

We assume that the learning processes are independent across doctors.[9] Therefore, we describe the learning process for doctor i, omitting the superscript i in most of the equations below for expositional clarity. We assume that, at time $t = 0$ (i.e., at *omeprazole*'s entry), she (doctor i) has the following *initial* beliefs about $\vec{\delta}$, the J-dimensional vector of quality differentials between *omeprazole* and the incumbent molecules:

$$\vec{\delta} \sim N\left(\begin{bmatrix} E_0\delta_1 \\ \vdots \\ E_0\delta_J \end{bmatrix}, \begin{bmatrix} \sigma^2_{\delta,1} & 0 & \dots & 0 \\ 0 & \sigma^2_{\delta,2} & \dots & 0 \\ 0 & \dots & \dots & \vdots \\ 0 & 0 & \dots & \sigma^2_{\delta,J} \end{bmatrix}\right). \tag{3.2.6}$$

All the information that she has about *omeprazole*'s efficacy in treating the various diagnoses should be summarized in the initial estimates $E_0\delta_1, \dots, E_0\delta_J$, while the initial variances $\sigma^2_{\delta,1}, \dots, \sigma^2_{\delta,J}$ parameterize her "confidence" in the accuracy of her initial estimates (higher confidence $\longleftrightarrow$ smaller initial variance).

How does the updating process work? Let's consider a generic time period t: assume that doctor i enters this period with beliefs that

$$\vec{\delta} \sim N\left(E_t\vec{\delta}, \Sigma_{\delta,t} \equiv E_t\vec{\delta}\vec{\delta}' - (E_t\vec{\delta})(E_t\vec{\delta})'\right) \tag{3.2.7}$$

During period t, the doctor prescribes *omeprazole* to k_j of her patients with diagnosis j, and observes k_j noisy signals of δ_j. We assume that the k_j signals (μ_{jtk}, $k = 1, \dots, k_j$) that she observes after prescribing *omeprazole* to k_j of her patients with diagnosis j take the following form:

$$\mu_{jtk} = \delta_j + \nu_{jtk} \tag{3.2.8}$$

[9] An interesting extension would be to introduce some kind of spillover *across* doctors.

where ν_{jtk} is normally distributed, with zero mean. In other words, each prescription yields a draw from a normal distribution centered around the true (but unknown) δ_j.

We assume that the correlation structure among the ν's for the different diagnoses are induced by a "one-factor" variance components structure for each ν:

$$\nu_{jtk} = \rho_j \theta_t + \eta_{jtk}, \; j = 1, \dots, J, \tag{3.2.9}$$

where:

- θ, the "common component" (across j and k) is distributed $N(0, \sigma_\theta^2)$, i.i.d. over t
- the η_{jtk}'s are independent over j, t, and k, and distributed $N(0, \sigma_{\eta j})$, $j = 1, \dots, J$
- $\rho_1, \dots, \rho_J$ are time-invariant parameters (the "factor-loadings").

This one-factor variance components specification reduces the number of parameters which we must estimate, while placing mild (and, we believe, acceptable) restrictions on the structure of correlation.[10]

Given these assumptions, then, the following correlation structure among all the signals (μ's) that she observes in period t emerges:

1. $\text{Var}(\nu_{jtk}, \; \nu_{jtk}) = \rho_j^2 \sigma_0^2 + \sigma_{\eta j}^2$
2. $\text{Var}(\nu_{jtk}, \; \nu_{jtk'}) = \rho_j^2 \sigma_0^2$, for $k \neq k'$
3. $\text{Cov}(\nu_{jtk}, \; \nu_{j'tk'}) = \rho_j \rho_{j'} \sigma_0^2$, for $j \neq j'$ and $\forall \; k, \; k'$.

For the rest of this section, let us assume that $J = 4$, and that in period t, two signals are observed for the first two diagnoses, and one each for diagnoses 3 and 4 (in other words, we observe $\vec{\mu}_t \equiv (\mu_{1t1}, \mu_{1t2}, \mu_{2t1}, \mu_{2t2}, \mu_{3t1}, \mu_{4t1})$). Also assume that, conditioning on all the

[10] One can extend the current model by allowing for a 2-factor variance components structure, but there may be problems identifying some of the parameters in that case.

signals received up to (but not including) period t, she believes that $\vec{\delta}$ is distributed:

$$\begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{pmatrix} \sim \left(\begin{bmatrix} E_t\delta_1 \\ E_t\delta_2 \\ E_t\delta_3 \\ E_t\delta_4 \end{bmatrix} \equiv E_t\vec{\delta}, \begin{bmatrix} \gamma_{1t}^2 & & & \\ \gamma_{12t} & \gamma_{2t}^2 & & \\ \gamma_{13t} & \gamma_{23t} & \gamma_{3t}^2 & \\ \gamma_{14t} & \gamma_{24t} & \gamma_{34t} & \gamma_{4t}^2 \end{bmatrix} \equiv \Sigma_{\delta t} \right). \tag{3.2.10}$$

Given these assumptions, then, in period t, doctor i believes that the vector of signals $\vec{\mu}_t$ has mean

$$\vec{\delta}_{\mu,t} \equiv \begin{bmatrix} E_t\delta_1 \\ E_t\delta_1 \\ E_t\delta_2 \\ E_t\delta_2 \\ E_t\delta_3 \\ E_t\delta_4 \end{bmatrix} \tag{3.2.11}$$

and variance-covariance matrix

$$\Sigma_{\mu,t} \equiv \begin{bmatrix} \gamma_{1t}^2+\rho_1^2\sigma_\theta^2+\sigma_{\eta 1}^2 & \gamma_{1t}^2+\rho_1^2\sigma_\theta^2 & \gamma_{12t}+\rho_1\rho_2\sigma_\theta^2 & \gamma_{12t}+\rho_1\rho_2\sigma_\theta^2 & \gamma_{13t}+\rho_1\rho_3\sigma_\theta^2 & \gamma_{14t}+\rho_1\rho_4\sigma_\theta^2 \\ & \gamma_{1t}^2+\rho_1^2\sigma_\theta^2+\sigma_{\eta 1}^2 & \gamma_{12t}+\rho_1\rho_2\sigma_\theta^2 & \gamma_{12t}+\rho_1\rho_2\sigma_\theta^2 & \gamma_{13t}+\rho_1\rho_3\sigma_\theta^2 & \gamma_{14t}+\rho_1\rho_4\sigma_\theta^2 \\ & & \gamma_{2t}^2+\rho_2^2\sigma_\theta^2+\sigma_{\eta 2}^2 & \gamma_{2t}^2+\rho_2^2\sigma_\theta^2 & \gamma_{23t}+\rho_2\rho_3\sigma_\theta^2 & \gamma_{24t}+\rho_2\rho_4\sigma_\theta^2 \\ & & & \gamma_{2t}^2+\rho_2^2\sigma_\theta^2+\sigma_{\eta 2}^2 & \gamma_{23t}+\rho_2\rho_3\sigma_\theta^2 & \gamma_{24t}+\rho_2\rho_4\sigma_\theta^2 \\ & & & & \gamma_{3t}^2+\rho_3^2\sigma_\theta^2+\sigma_{\eta 3}^2 & \gamma_{34t}+\rho_3\rho_4\sigma_\theta^2 \\ & & & & & \gamma_{4t}^2+\rho_4^2\sigma_\theta^2+\sigma_{\eta 4}^2 \end{bmatrix}. \tag{3.2.12}$$

Given the normality assumptions on the signals μ as well as on the δ's, her posterior beliefs about $\vec{\delta}$ given $\vec{\mu}_t$ is also described by a normal distribution whose mean and variance can be easily derived using the multivariate conditional mean and variance formulas. This "posterior distribution" in period t serves as the "prior distribution" for period $t+1$. In this way, we can derive the sequence of her posterior distributions over all the periods by repeatedly applying the conditional mean and variance formulas for jointly normally distributed random variables.

In order to do this, we must first characterize the joint distribution of $(\vec{\delta}, \vec{\mu}_t)$, during period t:

$$\begin{pmatrix} \vec{\delta} \\ \vec{\mu}_t \end{pmatrix} \sim N\left(\begin{bmatrix} E_t\vec{\delta} \\ \vec{\delta}_{\mu t} \end{bmatrix}, \begin{bmatrix} \Sigma_{\delta,t} & \Sigma_{\delta,\mu,t} \\ \Sigma'_{\delta,\mu,t} & \Sigma_{\mu,t} \end{bmatrix}\right) \tag{3.2.13}$$

where $\vec{\delta}_t$ and $\Sigma_{\delta,t}$ are, respectively, the mean and variance-covariance matrix of $\vec{\delta}$ conditional on all the signals received before period t. $\vec{\delta}_{\mu,t}$ and $\Sigma_{\mu,t}$ are the mean and variance-covariance matrix of the vector of signals $\vec{\mu}_t$, as given in equations 3.2.11 and 3.2.12, and $\Sigma_{\delta,\mu,t}$ is the matrix of covariance terms between $\vec{\delta}$ and $\vec{\mu}_t$ (which is easy to derive given equations 3.2.10, 3.2.8 and 3.2.9).

The first two moments of the posterior distribution of $(\vec{\delta} \mid \vec{\mu}_t)$ are

$$\vec{\delta}_{t+1} \equiv E_t(\vec{\delta} \mid \vec{\mu}_t) = \vec{\delta}_t + \Sigma'_{\delta,\mu,t}\Sigma^{-1}_{\mu,t}(\vec{\mu}_t - (\vec{\delta}_t)) \tag{3.2.14}$$

and

$$\Sigma_{\delta,t+1} \equiv \Sigma_t(\vec{\delta} \mid \vec{\mu}_t) = \Sigma_{\delta,t} - \Sigma'_{\delta,\mu,t}\Sigma^{-1}_{\mu,t}\Sigma_{\delta,\mu,t}. \tag{3.2.15}$$

Equation 3.2.14 yields the means of the posterior distribution of the δ's which are plugged into the logit prescription probabilities. These probabilities form the basis for our likelihood function, which is described in the next section.

Which parameters of the learning process do we estimate? First of all, we estimate the elements of the period zero initial mean vector $(E_0\delta_1, \ldots, E_0\delta_J)$ and the diagonal elements of the initial variance covariance matrix $(\sigma^2_{\delta,1}, \ldots, \sigma^2_{\delta,J})$. The true values $\delta_1, \ldots, \delta_J$ are parameters, as are $\rho_1, \ldots, \rho_J$ and $\sigma^2_{\eta 1}, \ldots, \sigma^2_{\eta J}$. We set σ^2_0 (the variance of the common component θ's) equal to 1; our attempts to estimate it have not been successful.

3.2.3 Remarks

In this section we raise several issues related to the learning model we have presented. First of all, the procedure we have outlined in the previous section to obtain a doctor's estimates of the δ's given her prescriptions during each period may seem cumbersome, especially given the much simpler posterior mean and variance formulas for sampling from a normal distribution with unknown mean and known variance: DeGroot ([28], chapter 9) gives formulas for both the univariate and multivariate case.[11] However, our problem fits into neither the univariate nor multivariate case. The difference arise because doctors don't observe complete vectors of draws from the multivariate distribution which they are trying to estimate the unknown mean of. To be precise, there are four diagnoses, but during some periods, doctors observe (say) two signals for diagnosis 1, 1 signal for diagnosis 3, and none for diagnoses 2 and 4. The theorems in DeGroot [28], chapter 9, are not suitable for this setting. Therefore, we sequentially derive the posterior means and variance-covariance matrices for the unknown δ's for each period t ($t = 1, \ldots, 31$) in the manner explained in the previous section.

Secondly, the rate of learning depends mostly on the diagonal elements — the variances — of the variance-covariance matrix of the signals (equation 3.2.12), which parameterize the "noisiness" of the signals. Roughly speaking, noisier signals (larger variances) slow down learning. What identifies the rate of learning in the data? Mainly *serial correlation* in *omeprazole* prescriptions over time, at the individual doctor level (the time dummies should soak up serial correlation at the aggregate level). A high degree of serial correlation at the individual doctor level implies "fast" learning, while absence of serial correlation implies slow learning. As such, one possible interpretation of the learning model is that it provides a structural explanation for serial correlation (conditional on covariates like *omeprazole*'s price and advertising) at the individual doctor level in prescriptions of *omeprazole* .

As such, we must be sure to control for other, simpler explanations of this micro-level

[11] A regression interpretation is also possible. Essentially, each doctor tries to estimate an unknown population mean using draws (or "signals") from a distribution centered at this mean, with known variance. Her "initial estimate" also constitutes an observation, but has a different variance (the "initial variance", parameterizing her confidence in her pre-sample information) than the other observations.

For the univariate case (i.e., one unknown population mean), the unbiased, minimum variance estimate of the population mean is the "GLS estimator", which is essentially a weighted average of the initial mean and the sample mean of the signals, where the weight given each observation is the inverse of the standard deviation (the "precision") of the distribution from which the observation is drawn. Analogously, for the multivariate case, where an estimate of a *vector* of unknown population means is desired using vectors of draws from a multivariate distribution centered around these means, the efficient estimator is a "joint-GLS" (seemingly unrelated regression) procedure.

serial correlation. The most obvious is *doctor heterogeneity.* To see this more clearly, assume that doctors know the true quality of *omeprazole*, but vary in their (time-invariant) propensity to prescribe it. What kind of prescription paths will emerge here? Doctors with high propensities will always prescribe a lot of *omeprazole* , while those with low propensities always a small amount — leading to a high degree of serial correlation in *omeprazole* prescription at the individual doctor level! Therefore in our specifications we make sure to control for both observed (using doctor-level covariates) and unobserved heterogeneity (using a random effects approach).

Thirdly, we allow for correlation among the signals only within a given time period t. As yet, we do not allow these signals to be correlated over time. Furthermore, by allowing for correlation among the disturbances to the signals (i.e., the ν's), the variance-covariance matrices describing a doctor's beliefs about $\vec{\delta}$ become non-diagonal, even though in period 0 (i.e., before *omeprazole*'s entry) this matrix is diagonal (see equation 3.2.6).

Fourthly, we assume that patients are perfectly homogeneous within a diagnosis cell. By doing this, we essentially assume perfect knowledge spillovers *within* cells, i.e., among patients with the same diagnosis. This assumption implies that if for example a new patient receives diagnosis j towards the end of the sample period, he can benefit from all the information the doctor has previously gathered by prescribing *omeprazole* to *any* of her previous patients. This specification assumes that — among all the patients receiving diagnosis j by doctor i at time t — the unobserved *iid* component drives whether or not a particular patient receives *omeprazole* . This is a common assumption in group data models. More general treatments would require multiple integration over patient heterogeneity distribution within each cell.[12]

Finally, doctors are *myopic* in our model — at any given time period, a doctor chooses a molecule based solely on the information that she has at that period, without taking into account the information that she would gain about *omeprazole* by prescribing it this period.

[12]On the other hand, since previous work on this dataset has shown the importance of patient level characteristics in the choice process, in future work, we would like to generalize the structural model by assuming a parametric distribution of patient heterogeneity within cells. (Some recent work by Berry, Levinsohn and Pakes [15] has explored ways of pooling micro and macro data in estimating demand models.) This would allow us to estimate knowledge spillovers across patients for the doctor.

As Shapiro [66], again, notes, with forward looking agents, products with uncertainty quality can be *more* attractive than products of known quality because there is no "down side risk" associated with the uncertain brand: "if it is a failure [the agent] simply returns to the old brand". Forward-looking decision making would involve early experimentation with new brands, so that the adoption path associated with a forward-looking learning model may not be adequate to explain the observed gradual rise in market share. Although (as discussed below) we see no evidence of such early experimentation in the data, a fully forward-looking learning model is nonetheless an interesting extension to the present model.

Ackerberg [1] has estimated such a forward-looking learning model. However, our learning model is more complicated because we permit not only the unknown quality differential parameter (δ) to vary across patients with different diagnoses (i.e., horizontal differentiation), but also the noisy signals observed by the doctors to be correlated across patients in a given month.

3.2.4 Econometric issues

In this section, we derive the likelihood function which we estimate. Given distributional assumption on the *i.i.d.* ϵ disturbances, the doctor's static expected utility maximization problem described earlier (3.2.3) yields choice probabilities for each observed prescription in the data. However, since we assume that, conditional on diagnosis, all the patients are homogeneous, we can aggregate up to the (doctor, month, diagnosis) level. Therefore a single observation in this model is how many patients of doctor i in the diagnostic cell j are prescribed *omeprazole* in month t (r^i_{jt}) out of the total number of patients with diagnosis j (N^i_{jt}) who are prescribed *any* anti-ulcer drug in month t by doctor i. [13]

Assuming that the ϵ's are distributed *i.i.d.* (over patients and time) extreme value, the likelihood function for a given observation takes the *grouped logit* form. Omitting the i

[13] This is equivalent to assuming that the doctor follows a random treatment strategy; that is, the doctor every month decides the proportion of her patients who will be treated with *omeprazole*, and randomizes among them.

superscript (and allowing $V_t^{i*} \equiv \alpha x^i + \beta \Delta p_t + \gamma \Delta \mathrm{ADV}_t + \xi_t$),

$$\mathrm{Prob}(r_{jt} \mid N_{jt}) = \binom{N_{jt}}{r_{jt}} \left(\frac{\exp(V_t^{i*} + E_t\delta_j)}{1 + \exp(V_t^{i*} + E_t\delta_j)} \right)^{r_{jt}} \left(\frac{1}{1 + \exp(V_t^{i*} + E_t\delta_j)} \right)^{N_{jt} - r_{jt}} \tag{3.2.16}$$

Second, the expected value of the "true" quality of *omeprazole* for diagnosis j in period t, $E_t\delta_j$, needs to be included in the grouped logit choice problem defined in equation (3.2.16). The main problem we encounter in the estimation process is that while the doctor observes the realization of the signal, μ_{jt}, we as econometricians don't. This implies that the likelihood function for a given sequence of prescription frequencies for a given (doctor-diagnosis type) combination involves an integral over the distribution of the unobserved signals:

$$\mathrm{Prob}(r_{11}, \ldots, r_{1T}, , \ldots, r_{J1}, \ldots, r_{JT} | N_{j1}, \ldots, N_{jT}, , \ldots, N_{J1}, \ldots, N_{JT}) =$$

$$\int \left[\prod_{t=1}^{T} \prod_{j} \binom{N_{jt}}{r_{jt}} \left(\frac{\exp(V_t^{i*} + E_t\delta_j)}{1 + \exp(V_t^{i*} + E_t\delta_j)} \right)^{r_{jt}} \times \left(\frac{1}{1 + \exp(V_t^{i*} + E_t\delta_j)} \right)^{N_{jt} - r_{jt}} \mid \vec{\mu} \right] dF(\vec{\mu}) \tag{3.2.17}$$

where $\vec{\mu}$ is the vector of signals observed by doctor i for patients with diagnosis j (but unobserved by the econometrician). Since a doctor observes one signal for each patient who she prescribes *omeprazole* to, the dimensionality of $\vec{\mu}$ is $R_T \equiv \sum_{j=1}^{J} \sum_{t=1}^{T} r_{jt}$, the number of prescriptions of *omeprazole* which the doctor wrote during the entire 31-month sample period to *all the* patients. In our data, R_T can run into the hundreds, making the integral intractable using traditional quadrature methods. We use simulation techniques (see McFadden and Ruud [53]) to evaluate these integrals.

We will estimate our model using simulated maximum likelihood. This procedure contains several steps. First of all, for each (doctor-diagnosis type) combination, we draw M sequences of disturbances with mean 0 and variance σ_j^2. Each sequence consists of R_T signals. Note that by definition (given above), R_T varies across doctors. In the results reported in this chapter, we used $M = 10$.

Secondly, for each sequence of signals, we calculate the 31 (one for each month) means $E_t\delta_j$, $t = 1, \ldots, 31$ using the standard normal conditional mean formula (reported in equation 3.2.14). Furthermore, the posterior mean of δ_j in month t will serve as the prior mean for month $t+1$. Having calculated the $E_t\delta_j$'s for $t = 1, \cdots, 31$ for a given drawn sequence of signals, we can calculate the grouped logit probabilities (in equation (3.2.16)) by substituting the calculated $E_t\delta_t$'s. Finally, we averaged the calculated grouped logit likelihood function (one for each drawn sequence) over all the drawn sequences.

Hence, the log-likelihood function for all (doctor-diagnosis type) observations is:

$$\begin{aligned}\text{LogL} = \sum_{i=1}^{326} \log \frac{1}{M} \sum_{m=1}^{M} \Bigg[\prod_{t=1}^{31} \prod_{j=1}^{J} \binom{N_{jt}}{r_{jt}} \left(\frac{\exp(V_t^{i*} + E_t\delta_j)}{1 + \exp(V_t^{i*} + E_t\delta_j)} \right)^{r_{jt}} \times \\ \left(\frac{1}{1 + \exp(V_t^{i*} + E_t\delta_j)} \right)^{N_{jt} - r_{jt}} \mid \vec{\mu^m} \Bigg]\end{aligned} \tag{3.2.18}$$

where 326 is the total number of doctors in the sample and J, the number of diagnosis groups, is 4 (note that the signals $\vec{v}$ are now indexed by m, which labels the drawn sequences).[14]

3.3 Data

Overview The data used in this analysis was collected by the Italian National Institute of Health. It records, for a 10% sample[15] of the doctors in the metropolitan area of Rome, all prescriptions of anti-ulcer drugs (therapeutic class A02B[16]) to all their patients during a three-year period (1990-1992). A prescription episode is the unit of observation in the dataset. The dataset contains more than 660,000 observations, each of which records the identity of the patient, the prescribing doctor, the drug prescribed, and the year and month of the prescription; 326 doctors, and 174,000 patients are represented in the data.

[14]Note that in the above equation we take averages of the simulated probabilities before taking the logarithm. It would have been much easier to average over the simulated log probabilities, since in that way we avoid the possible computational problems (specifically, numerical underflows) which can arise when taking the product over t and j of the prescription probabilities. The likelihood function that we maximize is $\sum_i \log E_{\vec{v}}\text{Prob}(p_{ijt} \mid N_{ijt})$. Simulation is necessary simply to calculate the expectation, which occurs *inside* the log operation. Since $\sum_i \log E_{\vec{v}}\text{Prob}(p_{ijt} \mid N_{ijt}) \neq \sum_i E_{\vec{v}} \log \text{Prob}(p_{ijt} \mid N_{ijt})$ the easier alternative (simulating the log probabilities) would be inappropriate.

[15]The doctors in the sample were not chosen following any sampling technique, since the only information available was their "id" number.

[16]This four digit code, the ATC code, is an international classification system according to which drugs are divided into different categories by target organ, mechanism of action, and chemical and therapeutic characteristics.

Figure 3.3 displays the distribution of the total number of prescription episodes of anti-ulcer drugs by the doctors during the three-year sample period. The median for the 326 doctors is around 2,000 prescriptions. Only a few doctors fill less than 1,000, or more than 3,500 prescriptions for anti-ulcer drugs during the sample period. Therefore, the doctor sample is fairly homogeneous in terms of practice size. There are two episodes of new molecule[17] entry in the Italian anti-ulcer drug market[18] during the sample period: *omeprazole* in June 1990, and *roxatidine* in January 1992. In this chapter we concentrate on entry of *omeprazole*, since *omeprazole* constituted a truly new treatment. *Roxatidine* is a "me-too" molecule (in the H_2-antagonist class) which has side-effects very similar to those of incumbent drugs. Moreover, in this study we do not distinguish among brand names for the same molecule (Coscelli [26] investigates competition among sellers of the same molecules.) We only focus on the *molecules* prescribed by the doctors.

Figure 3.4 shows the in-sample market shares during the three-year period for the *omeprazole*-based drugs and two large groups of competitors: H_2- receptor antagonists and "low-tech" drugs. The latter two groups of drugs constitute the non-*omeprazole* alternative (or "outside good") in our model.

Patient heterogeneity: different diagnoses As discussed near the beginning of this chapter, the anti-ulcer drug market is naturally segmented on the basis of diagnoses, so that a given molecule differs in its usefulness in treating different diagnoses (see figure 3.2). This consideration motivated the multivariate learning model presented in an earlier section, in which we permit *omeprazole*'s quality differential δ to vary across diagnoses. However, we do not observe a patient's actual diagnosis in the data: how can we allocate our observations to the four different diagnoses listed in figure 3.2?

We observe the quantity of anti-ulcer drugs prescribed, which varies according to the diagnosis: Therefore, after careful perusal of the medical literature, we decided to classify observations into the four diagnosis groups based on the *length* and *intensity* of treatments.

[17] A molecule is a clearly specified and usually patented chemical compound.

[18] The Italian pharmaceutical market is the fifth largest in the world. All the major drugs listed above are marketed there.

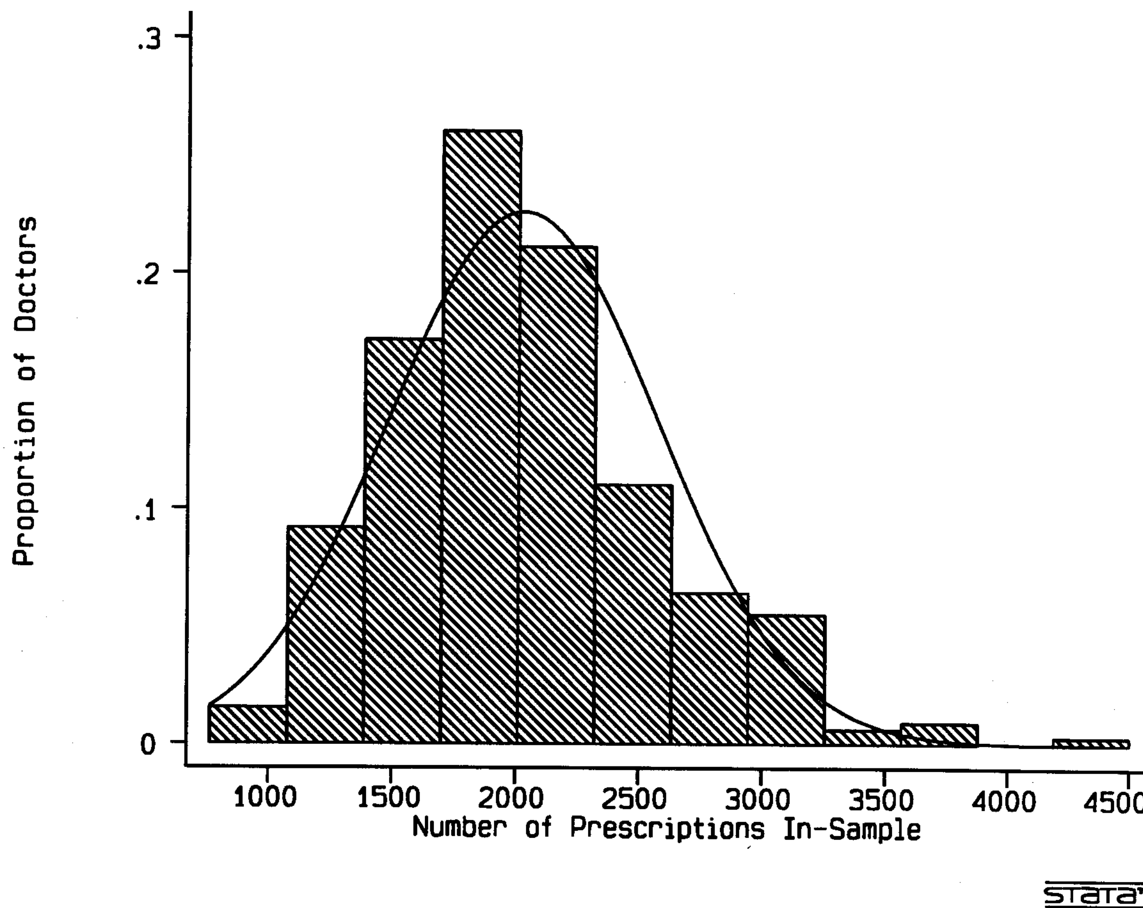


Figure 3.3: Total number of in-sample prescriptions by the 326 doctors. Sample period: t=1,...,36–January, 1990–December, 1992. **Source:** authors' computations using the sample data described in section 3.3.

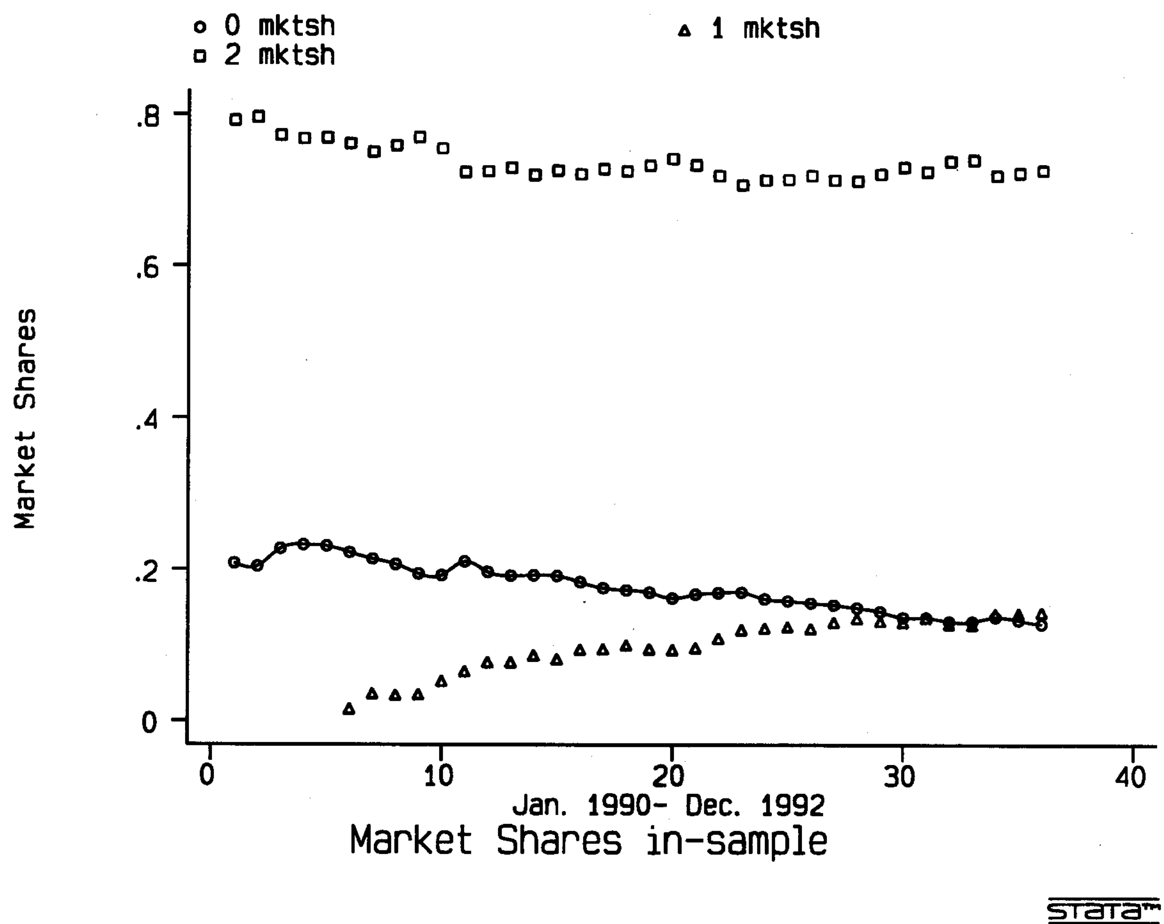


Figure 3.4: In-sample market shares: t=1,...,36–January 1990-December 1992. 0-"Low-tech" drugs, 1-*omeprazole* and 2-H_2- receptor antagonists. Entry of *omeprazole* took place at t=6–June, 1990. **Source:** authors' computations using the data described in section 3.3

Figure 3.5 describes the diagnoses, the possible treatments, the preferred drugs and the criteria we use to map prescriptions into diagnoses (Physicians' Desk Reference [56]).

Fist of all, from our data, we distinguish between three mutually-exclusive patient types. **Heartburn patients** are those with <= 2 in-sample prescriptions (regardless of the quantity prescribed); all the observations associated with these patients are classified as Diagnosis 1.

Patients with **pathological hypersecretory conditions** require large dosages of anti-ulcer medicine each month; we distinguish them in our data as those who are prescribed at least 133% of the average daily dose of anti-ulcer drugs in *every* month during which they are prescribed anti-ulcers drugs. The observations from these patients fall into diagnosis group 2.

Patients who are not identified as either of the first two types are defined as suffering from either **GERD or peptic ulcers.**[19] Observations for these patients can be classified into one of two diagnosis groups, corresponding to "attack" (diagnosis 3) and "maintenance" (diagnosis 4) therapy. For observations belonging to these patients, we distinguish among *patient-months*: if the amount of the drug prescribed in a month is greater or equal than 133% of the average daily dose for maintenance in that month, we classify the patient to diagnosis 3 (attack therapy) *for that month*; otherwise, the patient is classified to diagnosis 4. This is because during the sample period (i.e., before the advent of antibiotic treatments for duodenal ulcers), the usual treatment would consist of an attack therapy, followed either by no prescription or maintenance. If there was no maintenance therapy, the proportion of recurrences was very high (70-80% according to Graham [34]). A recurrence would be followed by a new attack therapy, and so on. Therefore, a typical patient suffering from either a peptic ulcer or GERD would cycle through periods of attack therapy, maintenance, no prescription, recurrence, attack, and so on — a regimen which we try to capture by our classification scheme.

[19]It would have been interesting to further distinguish between peptic ulcer and GERD, since *omeprazole* was proven in clinical trials to be more effective at treating GERD, but it is not possible to distinguish between the two in our data since the prescribed daily doses for a treatment regimen for either peptic ulcers or GERD is the same.

Figure 3.5 shows the proportion of the observations allocated to the 4 diagnostic classes. Cell 2 contains less than 5% of the prescriptions, while cell 4 contains more than 50% of the total number of prescriptions in the sample. These proportions are consistent with the incidence in the overall Italian population (peptic ulcer 59%, GERD 5%, minor heart-burns 23%, and "others" 12%) according to a survey of Italian doctors in 1991 cited in Pertile [62].

Figure 3.5: Observations in the 4 diagnostic cells

Diagnostic cell	**Empirical Distinction**[a]	**Frequency**	**Percent**
1–Minor heartburn	Patient has <= 2 in-sample prescriptions	135,466	20.49
2–Pathological hypersecretory conditions	$Q >= 133\%$ average monthly quantity for ulcer in *every* month	32,362	4.89
3–Attack therapy for ulcer or GERD	$Q >= 133\%$ average monthly quantity for ulcer in the month	132,361	20.02
4–Maintenance therapy for ulcer or GERD	$Q < 133\%$ average monthly quantity for ulcer in the month	361,054	54.60

Sample: N=661,243.
Source: data described in section 3.3.

[a]Prescription assigned to diagnoses by the authors using the daily dosage for the average patient requiring an ulcer treatment as suggested by *Physicians' Desk Reference* [56].

Next we describe how *omeprazole* diffused across the different diagnoses in our sample period. Figure 3.6 illustrates the market shares of *omeprazole*, H_2- receptor antagonists, and "low-tech" drugs in the four diagnostic cells over time. Although the trend is similar across the four cells. we can see that the market share of *omeprazole* (plotted in crosses) increased more for diagnoses 2 and 3, which are the ones where the clinical trials indicated that *omeprazole* was more effective.

Nonetheless, the large incidence of prescriptions for diagnosis 1 ("minor heartburns") evidenced in figure 3.5 as well as the rise in use of *omeprazole* in treating it (from figure 3.6) is troubling. It suggests that "heartburn" may be a mischaracterization of these patients' gastro-intestinal woes — after all, perhaps they suffer from something more serious than heartburns, given that they have decided to seek medical help rather than purchasing widely available over-the-counter heartburn medicines. As we discussed before, standard treatment for ulcers calls for a (short) period of intense "attack" therapy (using a powerful molecule such as *omeprazole*) followed either by a maintenance period or cessation of treatment. Diagnosis 1 may well encompass many patients who underwent attack therapy followed by cessation, for whom *omeprazole* would have been an appropriate molecule.

Furthermore, there may be a "survival bias" issue here: perhaps those patients whose treatment lasted only two months or less (and therefore classified in diagnosis 1) are exactly those for whom *omeprazole* had a high success rate. More generally, if *omeprazole* is a very effective molecule, leading to less frequent prescriptions and shorter treatment lengths, then our classification scheme can lead to selection bias because, essentially, inclusion in a particular diagnostic cell is endogenous. Therefore, below we also estimate and report the results from a model assuming patient homogeneity, without attempting to distinguish among diagnoses.[20]

From aggregate to individual level data Before estimating the model described in section 3.2.1, which is based on individual doctor data, we want to describe how heterogeneous doctors appear to be in their choice of *omeprazole* versus incumbent drugs.

Do doctors have different times to first adoption of *omeprazole*? Figure 3.7 shows when the doctors in our sample started prescribing *omeprazole* to at least one of their patients: almost 30% "adopted" in the first month; 60% adopted before the end of the second month and almost 75% prescribed *omeprazole* before the end of the third month. Finally, by the end of the 10th month *every* doctor had prescribed *omeprazole* to at least one patient. This constitutes preliminary evidence that the doctors are fairly homogeneous in their adoption decisions.

[20] In future work, we plan to explicitly model the effect of treatment option on treatment length ("success").

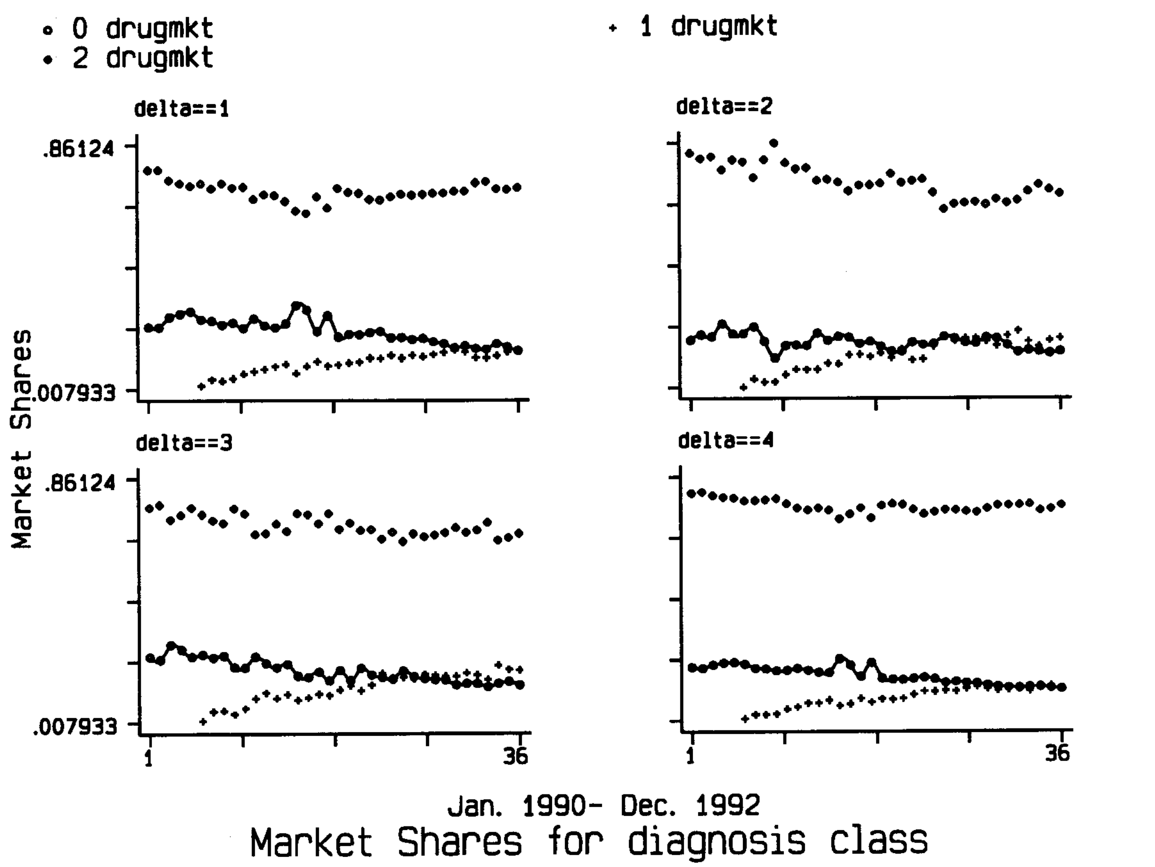


Figure 3.6: Market shares of *omeprazole* (drugmkt1), "low-tech" drugs (drugmkt0) and H_2-receptor antagonists (drugmkt2) in the *four* diagnostic cells over time. delta1—Minor Heart burns, delta2—Pathological Hypersecretory Conditions, delta3—Attack therapy for GERD or peptic ulcer, and delta4—Maintenance Therapy for GERD or peptic ulcer. t=1,...,36–January 1990 to December 1992. **Source:** authors' computations using the data described in section 3.3.

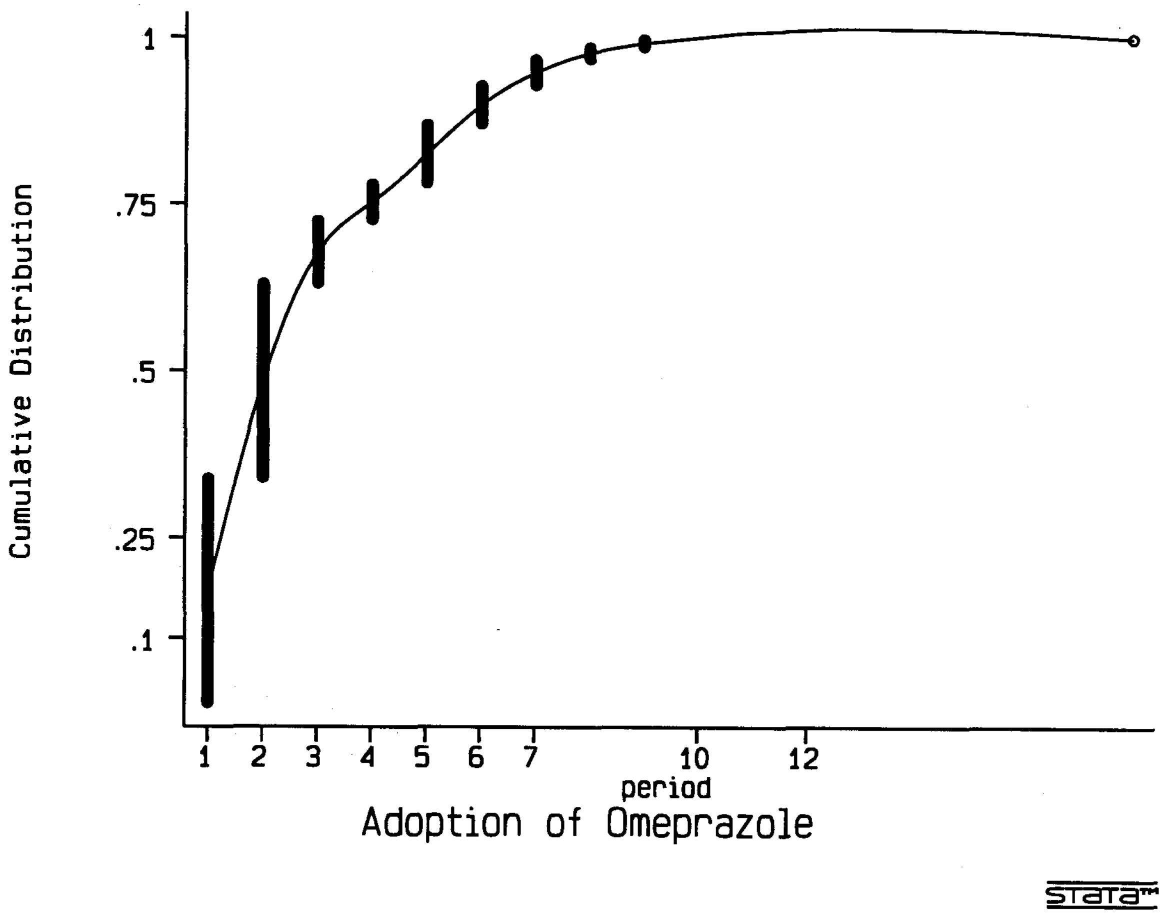


Figure 3.7: Time of adoption of *omeprazole* by the doctors in the sample. Vertical bars represent the doctors who adopted for the first time during that month. Analysis starts with entry of *omeprazole*: t=1–June 1990. t=1,...,31–June 1990-December 1992. Time of adoption is defined as the time of the first prescription of *omeprazole*. **Source:** authors' computations using the data described in section 3.3.

To further investigate doctors' heterogeneity, we analyze whether the doctors who immediately adopted *omeprazole* subsequently displayed a prescribing behavior different from the "late" adopters. We divide the doctors into three groups: early, medium and late adopters. Early adopters are the doctors who prescribed *omeprazole* both in the first and second month after entry (approximately 25% of the sample); late adopters are all the doctors who did not adopt *omeprazole* in the first three months (approximately 25% of the sample), while medium adopters are everyone else. Figure 3.8 plots the monthly market shares of *omeprazole* for the three groups (these are averages across all doctors who belong to the group). They never cross: early adopters have a constantly higher market share than the other two groups. The rate of growth in market shares is *very similar* across the three groups. This seems to indicate that some doctors adopted earlier, but then the diffusion of *omeprazole* across doctors was fairly homogeneous. This figure rules out theories of "immediate learning" whereby doctors learn the *omeprazole*'s true quality immediately upon making their first prescriptions.[21] This figure therefore shows that even after controlling for some degree of heterogeneity among doctors (namely for heterogeneity in preferences for *omeprazole* which would be reflected in different first adoption times), we still need to explain the rise in *omeprazole*'s market share over time.

While that provides some evidence in favor of the learning hypothesis,[22] it is far from a confirmation. Other factors not related to learning — such as *omeprazole*'s advertising over the sample period — could very well have caused the gradual rise in *omeprazole*'s market share. In order to measure the importance of learning, therefore, we must estimate a model in which not only learning from past prescriptions, but also advertising and marketing variables are allowed to influence a doctor's chances of prescribing *omeprazole*. Only by estimating such a model can we draw conclusions about the value of first-hand experience (as measured by doctors' actual prescriptions of the molecule) — and therefore the extent of informational barriers to entry — in this market.

[21] This is because otherwise we would observe the group market share for the *early adopters* jump to the asymptotic one immediately after the first prescription.

[22] $H0$: Doctors observe a noisy signal about the "true" quality of *omeprazole* when they prescribe it to their patients.

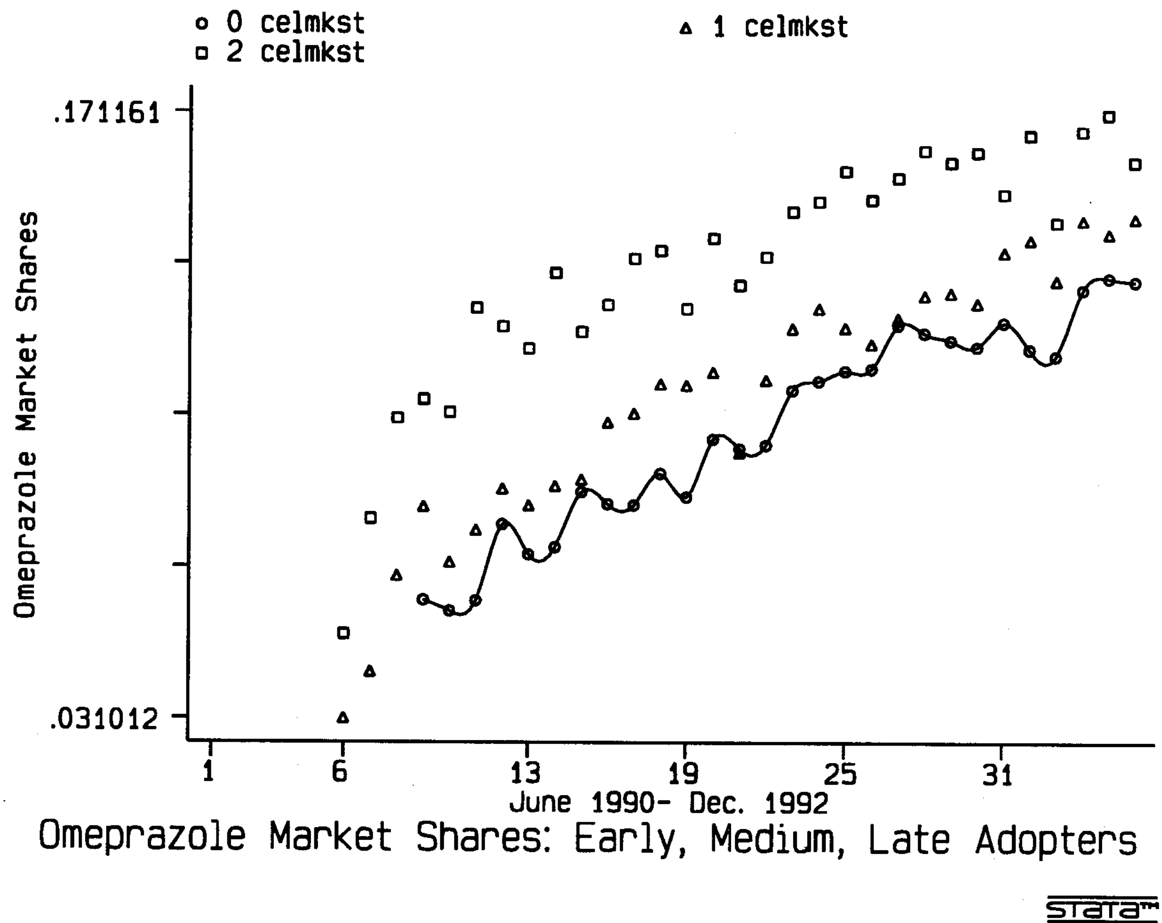


Figure 3.8: Market shares of *omeprazole*: monthly averages for early, medium and late adopters. 0-*late adopters* (those who did not adopt during the first three months), 1-*medium adopters* (those who do not belong to group 0 or 2), and 2-*early adopters* (those who adopted both during the first and second month after entry). Market shares are the monthly averages across all doctors belonging to a certain group. Entry took place in June 1990—t=6. **Source:** authors' computations using the data described in section 3.3

Finally, in our theoretical specification (section 3.2.1), we assume that there is no "forward-looking" component in the doctor's utility function. If such forward-looking behavior was significant we would expect to see many prescriptions of *omeprazole* clustered in the early months, because doctors would find it optimal to gather as much information as possible in the early months by experimenting. We do not observe this clustering in figure 3.8, which seems to indicate that any forward-looking motive is not of primary importance in the doctors' overall decision process.

3.3.1 Description of the covariates

Doctor heterogeneity We control for both observed and unobserved heterogeneity in doctors' propensities to prescribe a drug based on the *omeprazole* molecule.

Observed heterogeneity x^i includes three characteristics of doctor i (in equation 3.2.18 are the following, all of which control for observed factors which may affect doctor i's propensities to prescribe the new drugs based on *omeprazole*. They are: (i) the monthly average of Herfindahl index[23] at *brand* level; (ii) the monthly average of the Herfindahl index at the *molecule* level; and (iii) the average monthly quantity of anti-ulcer drugs prescribed. All of these are calculated based on doctors' observed prescription behavior *before omeprazole's entry*.

The first two variables — the two Herfindahl indices — control for unobserved factors (either due to the doctor or to her patient pool) which make her prescribe a variety of drugs. Previous work on the Italian pharmaceutical market (Coscelli [25]) shows that when the entering drug is identical or very similar to existing ones, the higher these variables are (so the fewer the number of drugs that a doctor tends to prescribe to her patients), the less likely the doctor is to prescribe a new drug to her patients. These findings might not hold here, though, because *omeprazole* is a truly innovative drug, therefore its therapeutic quality vis-a-vis the quality of the competing brands becomes a crucial factor in the choice

[23]The Herfindahl index is a standard measure of concentration. It is computed as the sum of the squares of the market shares. A high value of the Herfindahl index indicates that the doctors are concentrated in their prescriptions across brands.

process.

The third doctor-specific variable (average monthly quantity) controls for unobserved factors which make her prescribe more anti-ulcer drugs to her patients in general. We would expect that unconditionally doctors who prescribe a larger quantity of anti-ulcer drugs would be more likely to prescribe a new drug, but it is uncertain if this relationship holds when we condition in addition on doctors' willingness to prescribe across drugs using the two other doctor-specific variables described above.

Unobserved heterogeneity We also control for unobserved heterogeneity of the following form: κ^i — constant over all period t and specific to doctor i — which parameterizes doctor i's unobserved propensity to prescribe the *omeprazole*-based drugs over the incumbent drugs. This unobserved propensity may arise due to, for example, benefits that doctor i receives from manufacturer's of *omeprazole*-based drugs which are unobserved by the econometrician.

We assume that this effect shifts the intercept of doctor i net utility from prescribing *omeprazole* at time t. κ^i is modeled in *random effect* fashion; i.e., we specify a distribution for it (assumed identical over all doctors i) and integrate out over this distribution in evaluating the likelihood function. Due to the complexity of the likelihood function so far (equation 3.2.18) we restrict κ^i to have a 2-point distribution:

$$\kappa^i = \begin{cases} \kappa^1 & \text{with probability } q \\ \kappa^2 & \text{with probability } 1-q \end{cases} \quad \text{for all } i. \tag{3.3.19}$$

Furthermore, we normalize κ^i to have zero mean; i.e., $E\kappa^i = 0 \longleftrightarrow \kappa^2 = \dfrac{-q\kappa^1}{1-q}$, so the specification adds the two parameters (κ^1, q) to our model.

The resulting simulated log-likelihood function for the prescriptions written by doctor i takes a multinomial form:

$$\log l^i = q\log(l^{i,s} \mid \kappa^1) + (1-q)\log(l^{i,s} \mid \kappa^2) \tag{3.3.20}$$

where (similar to equation 3.2.18)

$$\log(l^{i,s} \mid \kappa) = \log \frac{1}{M} \sum_{m=1}^{M} \left[\prod_{t=1}^{31} \prod_{j=1}^{J} \binom{N_{jt}}{r_{jt}} \left(\frac{\exp(V_t^{i*} + E_t\delta_j + \kappa)}{1 + \exp(V_t^{i*} + E_t\delta_j + \kappa)} \right)^{r_{jt}} \times \left(\frac{1}{1 + \exp(V_t^{i*} + E_t\delta_j + \kappa)} \right)^{N_{jt}-r_{jt}} \mid \vec{\mu^m} \right] \tag{3.3.21}$$

Some of the results reported below are obtained by maximizing this likelihood function.

Advertising We focus on the number of office visits made on behalf of each anti-ulcer molecule as our measure of advertising for that molecule. This choice of number of visits, rather than other measures of marketing such as advertisements in journals or number of free samples distributed, was dictated to large extent by the data (from IMS Italy) — which for these other variables is spotty at best. Only for office visits do we have reasonably complete data over the sample period. We have observations on the number of office visits in Italy made on behalf of each molecule on a quarterly basis; for the purposes of this model, we have created variables which are the *ratio* of the number of visits made on behalf of *omeprazole* versus visits made on behalf of all other molecules.[24]

In addition, we decided to use a flow measure of advertising (ADV) as opposed to a stock measure. It could very well be that given the appropriate value for the decay parameter, we could construct an advertising stock time series which exactly mimics the aggregate market share (and thus predicts it perfectly). Presumably, however, advertising stocks will already be captured by the time effects $\xi_2, \ldots, \xi_{31}$. [25] Figure 3.9 plots stocks and flow measure of advertising for *omeprazole* over the sample period. Clearly, *omeprazole*'s entry

[24] From equation 3.2.5, the appropriate covariate to include in the utility specification should be $\Delta \mathrm{ADV}_t$, the *difference* in advertising between *omeprazole* and the incumbent molecules in period t. However, this latter covariate moves very erratically over time, with large swings within short intervals of time. We preferred the ratio because its variation was much less erratic, and also because it is unitless.

[25] Furthermore, the question of flow versus stock doesn't really matter given these time effects, since the advertising coefficient γ (and, for that matter, the price coefficient β) are accordingly identified from time period 1, in which the time effect is normalized to be 0. The stock vs. flow distinction would not matter for period 1, since in that period, they are both the same. Furthermore, the time dummies would also soak up any

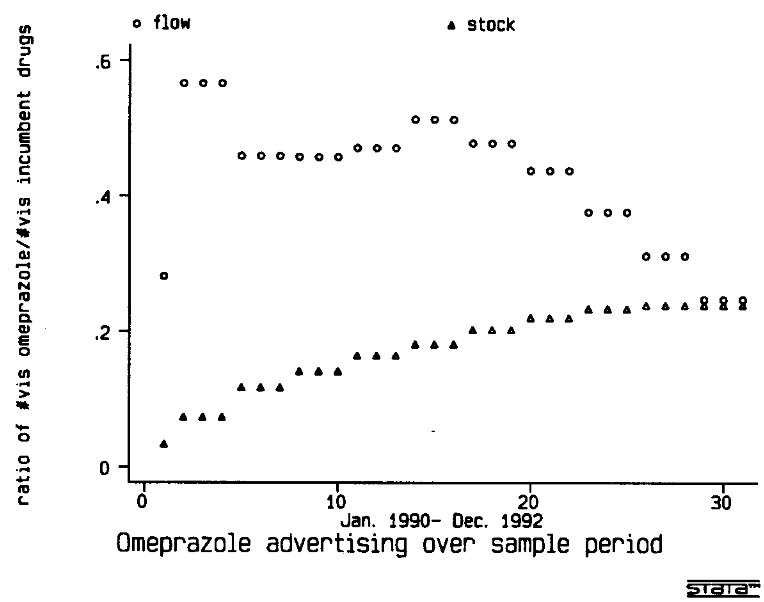


Figure 3.9: Advertising for *omeprazole* over the sample period
Note: advertising data is quarterly, so monthly graph resembles step function
In calculating advertising stocks, we used a depreciation factor of 0.8.[a]

[a]This follows on the previous work of BBRU [12] and Stern [68].

was accompanied by an aggressive marketing campaign, as evidenced by the high flow values in the earlier periods.

Summary statistics Figure 3.11 presents the summary statistics for the variables used in the estimates. The first block summarizes the number of observations we have for each (doctor, month) cell. There is an observation in a (doctor, month, diagnosis) cell if at least one patient receives the particular diagnosis by the doctor in that month. Therefore,

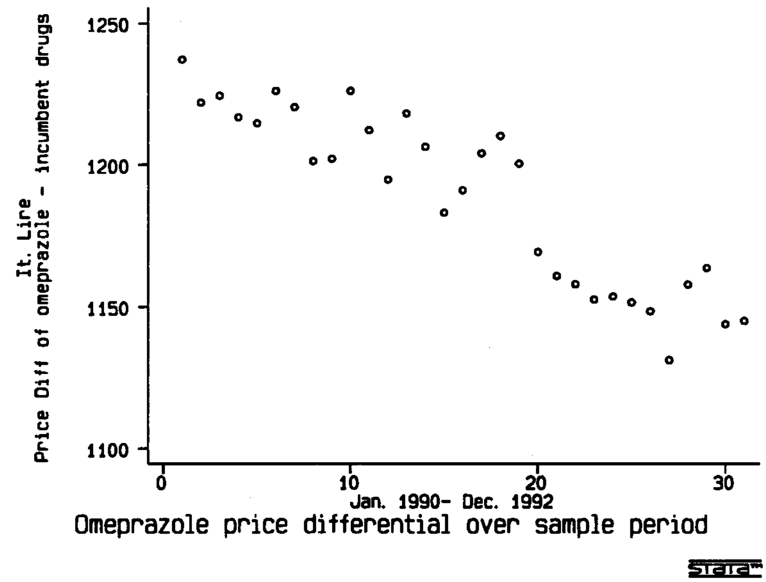


Figure 3.10: Price differential for *omeprazole* over the sample period

the upper bound on the number of observations we have for any diagnosis is 10,106 (326 doctors × 31 months). The least frequent is diagnosis 2: in only 5,567 (doctor-months) combinations at least one patient was diagnosed with a pathological hypersecretory condition. The column "Mean" indicates the average number of patients who respectively received a prescription of *any* anti-ulcer drug (the *total* row), or a prescription of *omeprazole* across the (doctor-month) cells *conditional on at least one* prescription of anti-ulcer drugs taking place. For example, for diagnosis 3 (attack therapy): on average there are 1.62 prescriptions of *omeprazole*, and 13.55 prescriptions of anti-ulcer drugs, conditional on at least one prescription of anti-ulcer drugs by doctor i in month s (8989/10106). The second block summarizes the values for the covariates proxying for doctors' heterogeneity, and the price difference of a day of therapy with *omeprazole* and with the outside good over time. *Omeprazole* had on average a 25% price mark-up over its competitors. The ratio of advertising of *omeprazole* over advertising for all the other anti-ulcer drugs available ranges between 24.7% and 56.5% during the sample period. Figure 3.10 plots this price difference over the sample period.[26]

3.4 Estimation results

3.4.1 Reduced-form results: models without learning

Before estimating the learning model, figure 3.12 reports the results of the estimation of a simple logit model where the decision to prescribe *omeprazole*, or the incumbent drugs depends solely on doctors' characteristics, price, and advertising. A time trend captures both first-hand experience, as well as second-hand sources of information (such as word of mouth). This model constitutes a very useful benchmark, because it is already more detailed than most of the models estimated in the literature.

The three time-invariant doctors' covariates are statistically significant; doctors who prescribe more anti-ulcer drugs, and who were more dispersed (i.e., high Herfindahl index) across anti-ulcer drugs in their prescription patterns prior to entry are more likely to prescribe *omeprazole*. These results are consistent with the findings for a different sample of Italian doctors (Coscelli [25]). Moreover, when the ratio of advertising for the incumbent drugs increases relative to the advertising of *omeprazole*, the likelihood of prescribing omeprazole decreases; finally, there is a positive time trend. The second specification uses the total quantity of *omeprazole* prescribed in the Italian market as a proxy for word of

[26] Over the sample period, *omeprazole*'s price differential decreases, but since prices are tightly regulated in Italy, this decrease is due less to price decreases in the non-*omeprazole* molecules, than to shifts in the composition of the non-*omeprazole* "basket" towards cheaper therapeutic alternatives.

Figure 3.11: **Summary Statistics**

	4 diagnoses cells (N=10,106: DOCTORS=326*31=T)				
Dependent variables–number of prescriptions in each (doctor,month) cell					
Variable	**# Obs.**	**Mean**	**St.Dev.**	**MIN**	**MAX**
MINOR HEART-BURNS					
omeprazole	10015	1.157	1.747	0	17
Total	10015	11.466	6.86	1	54
HYPERSECRETORY CONDITIONS					
omeprazole	5567	.706	1.613	0	26
Total	5567	5.342	6.373	1	70
ATTACK THERAPY					
omeprazole	8989	1.62	2.934	0	42
Total	8989	13.557	17.8	1	148
MAINTENANCE THERAPY					
omeprazole	10092	2.748	3.092	0	33
Total	10092	30.651	12.94	1	101
Doctors' Characteristics					
Herfindahl Brand level	326	.252	.068	.13	.564
Herfindahl Molecule level	326	.503	.113	.232	.805
Monthly quantity	326	798.37[a]	234.15	241	1682.08
Choice Characteristics					
Price diff.	31	1,188.724[b]	30.661	1131.17	1237.15
Ratio of ADV of *omeprazole*[c] *over all* the incumbent drugs	31	.427	.094	.247	.565

[a] The unit of measurement is ***defined daily doses*** (days of therapy).

[b] Italian Lire; 1US$=1,600 ItL; therefore, the average price difference between a day of treatment with *omeprazole*, and a day of treatment with the outside good is about 75 cents (this is approximately 25% of the price of a day of treatment with *omeprazole*).

[c] This data was provided by IMS Italy; the measurement unit is number of visits by the "detailers" during the quarter.

Figure 3.12: Simple Logit Specification

	Estimates	s.e.	Estimates	s.e.
Dependent variable: 1 if *omeprazole* is prescribed, 0 otherwise				
Doctors' characteristics				
HERFINDAHL PRODUCT LEVEL	-.636	.09	-.636	.09
HERFINDAHL MOLECULE LEVEL	-.519	.05	-.519	.05
QUANTITY/100	.016	.001	.016	.001
Choice characteristics				
Δp: $(\text{PRICE}_{omep} - \text{PRICE}_{og})/1000$	.002	.0003	-.002	.0002
ADV FLOW: OUTSIDE GOOD/*omeprazole*$^{(a)}$	-.264	.009	-.211	.009
TIME TREND	.065	.001		
QUANTITY PRES. IN THE NAT. MKT$^{(a)}$			.003	.00008
Constant	-6.231	.43	.528	.337
Log-likelihood fxn	-183750.6		-183890.87	

Number of observations: 574,314; 31 months out of the 36 in-sample months are included, since entry of *omeprazole* took place in June 1990.

(a) **Source:** authors' computations using data from IMS Italy.

mouth and diffusion of information about *omeprazole* instead of a simple time trend. The results are very similar between the two specifications; only the coefficient on the price difference changes. Since the price difference does not change much during the sample period, it is not very well identified when a constant is introduced into the estimation, therefore its sign changes in the two specifications, while the sign of the constant changes dramatically as well.[27] These results show that doctors are heterogeneous, and that their pre-entry dispersion across brands in their prescription patterns is a good predictor of the likelihood of prescribing a new drug. Moreover, a positive time trend and advertising flows help explain the observed changes in market shares. From this simple model, we move to the learning model, which explicitly accommodates learning via first-hand experience.

3.4.2 Results from structural learning model

Figures 3.13 and 3.14 reports the results obtained by estimating equation (3.2.18). Several specifications were estimated. Figure 3.13 contains results for a 1-Diagnosis specification which assumes no differences in δ across patient diagnosis groups. In such a model, all the signals which a doctor observes about *omeprazole*'s quality are *i.i.d.* draws from a distribution with mean δ and variance σ_η^2. Figure 3.14 contains two specifications of the 4-Diagnosis model described above, where we have categorized the various patient-visit observations into four diagnosis cells and the noisy signals of *omeprazole*'s quality to be drawn from four different distributions, characterized by diagnosis-specific mean (δ_i, $i = 1, \dots, 4$) and variance ($\sigma_{\eta i}$, $i = 1, \dots, 4$). We estimate a No-Spillover model in which signals are not allowed to be correlated across diagnoses, and a Spillover model where this correlation is allowed to be nonzero. The 1-Diagnosis model is estimated to gauge the statistical superiority of the 4-Diagnosis model described in the main text, which however is prone to the criticism that the criteria whereby we group patients in diagnosis cells may

[27]Price endogeneity will not be a major concern for us; in Italy, pharmaceutical prices are fixed by a central regulatory authority, and prices can be endogenous only to the extent that they are based upon some "unobserved quality" of the drug which doctors are aware of (but not the econometrician) and also base their prescription choices upon. Furthermore, if a time-invariant "unobserved quality" effect is present, it would be controlled for by the initial mean parameters which we estimate: $E_0\delta_1$, $E_0\delta_2$, $E_0\delta_3$, $E_0\delta_4$

Figure 3.13: Results: 1-Diagnosis Model
Number of observations: 10,106 (326 doctors/31 months)
Dependent variable: $r^i_{jt} = k$ if doctor i prescribes *omeprazole* k times to patients during month t, when her total number of anti-ulcer prescriptions to patients with diagnosis j is N^i_{jt}

	1-Diagnosis model	
	estimates	s.e.
Choice characteristics		
Δp: $(\text{PRICE}_{omep} - \text{PRICE}_{og})/1000$	-0.017	0.032
ADVERTISING (FLOW)	2.287	0.088
Doctors' characteristics		
HERFINDAHL PRODUCT LEVEL	0.225	0.021
HERFINDAHL MOLECULE LEVEL	-0.395	0.013
QUANTITY/100	-0.033	0.0004
Learning process parameters		
Initial Distribution		
INITIAL MEAN $E_0\delta$	-4.904	1.402
LOG(INITIAL ST.DEV.)	-4.301	0.361
Other parameters		
True quality differential – δ	-1.447	0.036
LOG(ST.DEV. OF SIGNALS)	-2.476	0.360
TIME PERIOD DUMMIES	yes	
Log-likelihood fxn	-55246.71	
M (# sim. draws)	10	

Figure 3.14: Results for 4-Diagnosis models
Number of observations: 10,106 (326 doctors/31 months)
Dependent variable: $r^i_{jt} = k$ if doctor i prescribes *omeprazole* k times to patients with diagnosis j during month t, when her total number of anti-ulcer prescriptions to patients with diagnosis j is N^i_{jt}

	No Spillovers		*W/ Spillovers*	
	estimates	s.e.	estimates	s.e.
Choice characteristics				
Δp: $(\text{PRICE}_{omep} - \text{PRICE}_{og})/1000$	-0.033	0.092	-0.021	0.028
ADVERTISING (FLOW)	1.957	0.247	1.890	0.064
Doctors' characteristics				
HERFINDAHL PRODUCT LEVEL	0.009	0.015	0.009	0.003
HERFINDAHL MOLECULE LEVEL	-0.361	0.013	-0.349	0.012
QUANTITY/100	-0.027	0.0004	-0.033	0.001
Learning process parameters				
Initial Distribution				
INITIAL MEAN – DIAG. 1 $E_0\delta_1$	-3.917	0.095	-3.391	0.041
INITIAL MEAN – DIAG. 2 $E_0\delta_2$	-3.293	0.098	-3.226	0.033
INITIAL MEAN – DIAG. 3 $E_0\delta_3$	-3.662	0.094	-3.550	0.037
INITIAL MEAN – DIAG. 4 $E_0\delta_4$	-4.376	0.094	-4.242	0.034
INITIAL VAR-COV MATRIX	see text			
Other parameters				
True quality diff.: diag. 1 (δ_1)	-1.097	0.098	-0.898	0.034
True quality diff.: diag. 2 (δ_2)	-0.041	0.096	-0.047	0.024
True quality diff.: diag. 3 (δ_3)	-0.975	0.096	-0.432	0.038
True quality diff.: diag. 4 (δ_4)	-1.384	0.091	-1.815	0.039
VAR-COV MATRIX OF SIGNALS	see text			
TIME PERIOD DUMMIES	yes		yes	
Log-likelihood fxn	-54700.59		-54343.49	
M (# sim. draws)	10		10	

Figure 3.15: Results for 4-Diagnosis model with unobserved doctor heterogeneity
Number of observations: 10,106 (326 doctors/31 months)
Dependent variable: $r^i_{jt} = k$ if doctor i prescribes *omeprazole* k times to patients with diagnosis j during month t, when her total number of anti-ulcer prescriptions to patients with diagnosis j is N^i_{jt}

	W/ Spillovers	
	estimates	s.e.
Choice characteristics		
Δp: $(\text{PRICE}_{omep} - \text{PRICE}_{og})/1000$	-0.010	0.420
ADVERTISING (FLOW)	1.025	0.861
Doctors' characteristics		
HERFINDAHL PRODUCT LEVEL	0.009	0.036
HERFINDAHL MOLECULE LEVEL	-0.728	0.025
QUANTITY/100	-0.009	0.001
Learning process parameters		
Initial Distribution		
INITIAL MEAN – DIAG. 1 $E_0\delta_1$	-3.497	0.257
INITIAL MEAN – DIAG. 2 $E_0\delta_2$	-2.784	0.253
INITIAL MEAN – DIAG. 3 $E_0\delta_3$	-3.123	0.253
INITIAL MEAN – DIAG. 4 $E_0\delta_4$	-3.903	0.249
INITIAL VAR-COV MATRIX	see text	
Other parameters		
True quality diff.: diag. 1 (δ_1)	-1.148	0.254
True quality diff.: diag. 2 (δ_2)	-0.086	0.253
True quality diff.: diag. 3 (δ_3)	-0.839	0.251
True quality diff.: diag. 4 (δ_4)	-1.846	0.254
VAR-COV MATRIX OF SIGNALS	see text	
Unobs'd Het. Dist'n Parms.		
κ^1	-0.223	0.016
q	0.614	0.066
TIME PERIOD DUMMIES	yes	
Log-likelihood fxn	-52888.52	
M (# sim. draws)	10	

be endogenous.[28] Finally, we also obtain estimates from the Spillovers model with doctor unobserved heterogeneity by maximizing the likelihood function from equation 3.3.20. These estimates are reported in figure 3.15.

Figure 3.16 contains the estimated initial (i.e., before *omeprazole*'s entry) variance-covariance matrices for the spillovers and no spillovers models, while figure 3.17 contains the estimated variance-covariance matrices for a hypothetical vector of disturbances in the signals (the ν_{jtk}'s) for a time period t.

On the basis of the likelihood function values alone, the Spillover model with doctor heterogeneity fits the data far better than either the Spillover model without heterogeneity, the No-Spillover model, or the 1-Diagnosis model. However, the 1-Diagnosis model was estimated mainly to investigate the extent of endogeneity in the diagnostic cell classifications, which cannot be gauged by comparing likelihood function values. However, the estimates from the 1-Diagnosis and 4-Diagnosis models vary little qualitatively and quantitatively, thus offering some indirect evidence that this type of endogeneity is not a big problem here.

The learning parameters According to all the sets of results, doctors initially assume that the basket of incumbent drugs is better than *omeprazole* in treating every diagnosis (we estimate the initial mean of the quality differential between *omeprazole* and the basket of incumbent drugs to be large and negative).[29] Our estimates of the δ's, (which are allowed to differ across diagnoses in the 4-Diagnosis models) are also all negative but, for all the specifications, much higher than the initial means, indicating that, relative to their beliefs later in the sample period, doctors are very pessimistic about *omeprazole*'s quality at the time of its entry.

This finding confirms the existence of informational barriers to entry which drive *omeprazole*'s slow acceptance into the anti-ulcer market. Note that uncertainty by itself is not

[28]Endogeneity will be a problem if drug quality is correlated with treatment length, since one way we categorize patients in diagnosis cells is by length of treatment.

[29]We assume that *all* doctors have the same initial beliefs about $\vec{\delta}$, *omeprazole*'s "true" quality differential versus the incumbent drugs. This assumption is consistent with the idea that doctors' initial beliefs are shaped by advertising taking place before entry, and publications in medical journals.

Figure 3.16: Estimated Initial Variance-Covariance Matrices

- For No-Spillovers model

$$10^{-3} * \begin{bmatrix} 0.2806 & 0 & 0 & 0 \\ 0 & 0.1663 & 0 & 0 \\ 0 & 0 & 0.2314 & 0 \\ 0 & 0 & 0 & 0.2463 \end{bmatrix}$$

- For Spillovers model:

$$10^{-3} * \begin{bmatrix} 0.4335 & 0 & 0 & 0 \\ 0 & 0.2276 & 0 & 0 \\ 0 & 0 & 0.3588 & 0 \\ 0 & 0 & 0 & 0.5064 \end{bmatrix}$$

- For Spillovers model with 2-point unobserved doctor heterogeneity:

$$10^{-3} * \begin{bmatrix} 0.1522 & 0 & 0 & 0 \\ 0 & 0.1679 & 0 & 0 \\ 0 & 0 & 0.1250 & 0 \\ 0 & 0 & 0 & 0.1733 \end{bmatrix}$$

enough to lead to informational barriers to entry — we require also, as in this case, initial estimates of the new product's quality which are biased downwards (pessimistic). As Shapiro [66] pointed out, if agents choose the product yielding highest *expected* utility (as our doctors do here), uncertainty about product quality, by itself, can cause a new product's expected utility to be higher than an incumbent's (known) utility, thus actually *facilitating* entry. Without something else which shifts down agents' *ex ante* utility from trying the new product — pessimism in our analysis here, and "brand loyalty" in Shapiro's case — informational barriers to entry will not arise.[30]

This finding of pessimism is, however, not unexpected. As we noted previously (in section 3.2.3), the parameters of the learning model are most identified off serial correlation in prescriptions of *omeprazole* at the individual doctor level. Since the data exhibits very strong serial correlation in prescriptions at the doctor level, this is rationalized in the

[30]Let's quote Shapiro again: "Basically, the argument that entrants are penalized because of subjective uncertainty about their products is wrong, although they are penalized *if their products are on average expected to be worse than established brands*" ([66], pp. 5-6, our emphasis.)

Figure 3.17: Estimated Var-Covariance Matrices for the Signals
Hypothetical example: Two signals each for diagnoses 1 and 2; one signal each for diagnoses 3 and 4

- For No-Spillovers model

$$\begin{bmatrix} 0.0036 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0036 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0018 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0018 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0031 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0047 \end{bmatrix}$$

- For Spillovers model:

$$\begin{bmatrix} 0.0053 & 0.0011 & 0.0001 & 0.0001 & 0.0013 & -0.0020 \\ 0.0011 & 0.0053 & 0.0001 & 0.0001 & 0.0013 & -0.0020 \\ 0.0001 & 0.0001 & 0.0022 & 0.0000 & 0.0001 & -0.0002 \\ 0.0001 & 0.0001 & 0.0000 & 0.0022 & 0.0001 & -0.0002 \\ 0.0013 & 0.0013 & 0.0001 & 0.0001 & 0.0053 & -0.0024 \\ -0.0020 & -0.0020 & -0.0002 & -0.0002 & -0.0024 & 0.0097 \end{bmatrix}$$

- For Spillovers model with 2-point unobserved doctor heterogeneity:

$$\begin{bmatrix} 0.0012 & 0.0002 & 0.0001 & 0.0001 & 0.0003 & -0.0004 \\ 0.0002 & 0.0012 & 0.0001 & 0.0001 & 0.0003 & -0.0004 \\ 0.0001 & 0.0001 & 0.0014 & 0.0000 & 0.0001 & -0.0002 \\ 0.0001 & 0.0001 & 0.0000 & 0.0014 & 0.0001 & -0.0002 \\ 0.0003 & 0.0003 & 0.0001 & 0.0001 & 0.0010 & -0.0006 \\ -0.0004 & -0.0004 & -0.0002 & -0.0002 & -0.0006 & 0.0018 \end{bmatrix}$$

framework of the learning model as initial pessimism.

Nonetheless, some noteworthy trends emerge in comparing the *degree* of pessimism among the four different diagnoses, as estimated in the 4-Diagnosis models. The spillover parameters (not reported in figure 3.14) are small in magnitude (as demonstrated by the small off-diagonal elements in the second matrix in figure 3.17), and not very precisely estimated. Therefore it is not surprising that the results for both the No-Spillovers and Spillovers models are very similar.

The only way they differ is in what the estimates of the initial means imply about doctors' initial ranking of *omeprazole*'s efficacy relative to the basket of incumbent drugs for the four different diagnoses. For the No-Spillover model results, the doctors start with an initial ranking of (from highest to lowest efficacy): 2,3,1, and 4. For the Spillovers results, this ranking is 2,1,3,4 (so upon *omeprazole*'s entry, the Spillovers results differ from the No-Spillover results in implying that doctors believe *omeprazole* to more more effective in treating heartburns (diagnosis 1) than attacking peptic ulcers (diagnosis 3)). However, the "true" ranking (from the estimates of the true quality differentials) is the same in both models: 2,3,1,4; this agrees with the initial ranking in the No-Spillover results.

These estimates indicate that *omeprazole* is mostly beneficial (relative to the basket of incumbent drugs) in treating pathological hypersecretory conditions (diagnosis 2), which is a reasonable conclusion. For the Spillovers results, the "reversal" in rankings between diagnoses 1 and 3 imply that early on, doctors overestimate the *omeprazole* 's usefulness for heartburns (diagnosis 1) but underestimate it for "attacking" peptic ulcers (diagnosis 3).

We've focused on ranking and signs of these coefficients because the actual magnitudes of the estimated initial means and "true" quality differentials are difficult to interpret, as they depend highly on our choice of the other covariates to include in the utility specification. Later on, we will present some computational results which illustrate the extent to which doctors update their beliefs concerning *omeprazole*'s quality differential, as implied by these estimates.

As noted earlier, the initial variance-covariance matrices parameterize a doctor's confidence in her initial estimates regarding *omeprazole*'s quality differential relative to the incumbent drugs. The magnitudes in these matrices, calculated in figure 3.16, are extremely small (roughly three orders of magnitude smaller than the estimates of the initial means), implying a great deal of confidence in these initial (and, as we saw above, pessimistic) estimates. For both the Spillover and No-Spillover specifications, we find that doctors are "most confident" about their initial estimate of *omeprazole*'s quality for diagnosis 2, pathological hypersecretory conditions (for the Spillovers model, the initial variance (0.2276) is lowest for this diagnosis). After controlling for time-invariant unobserved heterogeneity at the doctor level, however, the smallest element in the initial variance-covariance matrix (the bottom one in figure 3.16) is for diagnosis 3, "attack" therapy. Neither of these findings are surprising since, as discussed before, the medical evidence seemed unambiguously clear about *omeprazole*'s effectiveness in treating these two diagnoses. On the other hand, the finding that doctors are very confident in their pessimistic estimates is troubling: if we were to find, in addition, that the signals which doctors receive are drawn from distribution with *large* variances (i.e., "noisy" signals), the learning process would be slow, since the initial confidence will exacerbate the downward bias in doctors' subsequent updated estimates of the δ's.

Fortunately, this is not true. The magnitudes of the diagonal elements of the estimated variance-covariance matrices of the signals (reported in figure 3.17) are also small, similarly implying that the signals are very precise. Again, for the 1-Diagnosis and Spillovers model without unobserved doctor heterogeneity, the signals received from patients with pathological hypersecretory conditions (diagnosis 2) are the most precise (their variance — 0.018 for the No-Spillover results, and 0.0022 for the Spillovers results are the lowest among the four diagnoses). Mimicking the trend observed in the initial variance-covariance matrices discussed in the previous paragraph, the results from the Spillovers model after accounting for unobserved doctor heterogeneity differ in indicating that the signals from diagnosis 3 patients — those undergoing "attack" therapy, according to our classification scheme — are the most precise.

Furthermore, estimates of the off-diagonal elements for the variance-covariance matrices for the signals from both of the Spillovers models (both with and without unobserved doctor heterogeneity) indicate mostly positive spillovers, except for the signals associated with diagnosis 4 ("maintenance" therapy). The negative values for the correlation terms involving this diagnosis indicate that a positive outcome in prescribing *omeprazole* to a patient for diagnoses 1–3 leads doctors to deem *omeprazole* is less attractive for maintenance purposes — not surprising, since (as described before) maintenance therapy typically requires longer than the eight weeks which is the appropriate length of a treatment involving *omeprazole*. This confirms the last column of figure 3.2 (derived independently from medical sources) that the anti-ulcer market is *horizontally differentiated*: since a doctor will rank the two alternatives (*omeprazole*-based drug versus incumbent drugs) differently depending on the diagnosis of the patient who she's prescribing to.

Other parameters Recall that these results were obtained conditional on other covariates which could similarly predict the gradual rise in *omeprazole*'s market share over time. These covariates are: advertising (measured in terms of the ratio of period *flows* of doctor visits undertaken on behalf of *omeprazole*, versus those taken on behalf of the incumbent drugs), the price differential of *omeprazole* versus the incumbent drugs and, finally, 31 time-period (but not diagnosis-) specific dummies to soak up all the residual factors affecting the rise in *omeprazole*'s market share: these factors could include journal articles, conferences taking place after entry or word-of-mouth spillovers across doctors.

The coefficient on advertising is positive and significant in all the models (the point estimate is 1.890 for the Spillovers model), implying that doctors are more likely to prescribe *omeprazole* in those quarters when Italian detailing activity on *omeprazole*'s behalf is relatively intense. This confirms BBRU's [12] conclusions that marketing was key in ensuring the success of *Zantac* in the anti-ulcer market (where it still dominated during the sample period). The question then arises: how effective is advertising vis-a-vis direct first-hand experience in convincing doctors to prescribe *omeprazole*? This will be explored further below.

The coefficient on Δp, the price differential between *omeprazole* and all the other anti-ulcer molecules, is negative but not significant across all three models (-0.021 in the Spillovers model). This finding is not unexpected, since the Italian health care institutions (all medications is paid for by the National Health Service) do not force doctors to internalize the costs of the drugs that they prescribe. Furthermore, there is not much price variation in the sample, and the price of the alternative drugs is a weighted average of the prices of *all* the incumbent drugs.

Finally we turn to the doctor-specific covariates which capture observed time-invariant heterogeneity across doctors in their willingness to prescribe *omeprazole*. The two covariates which measure dispersion in prescription patterns prior to *omeprazole*'s entry are generally statistically significant[31] but have different signs: we find that greater dispersion at the *molecular* level (lower value for the molecular-level Herfindahl index) increases the probability of prescribing *omeprazole*, but not greater dispersion at the product level. Furthermore, larger quantity prescribed is associated with *lower* propensity to prescribe *omeprazole*. These latter two effects are opposite in sign from those we obtained for the simple logit specification in figure 3.12. Clearly these three covariates will be quite correlated at the doctor-level, so that these changes in signs across different specifications is not surprising. On the other hand, the results described here are robust across all four specifications of the learning model (the 1-Diagnosis, and all three 4-Diagnosis models); furthermore, if we interpret the learning model as a way of controlling for unobserved doctor- and time period specific effects, then these estimates are also robust to this type of unobserved heterogeneity, which is not true for the estimates obtained in the simple logit specifications, which assume that prescriptions are *i.i.d.* over doctors and time period conditional on the observed covariates.

Importance of first-hand experience Our estimates have allowed us to conclude that at the time of *omeprazole*'s entry, doctors were very pessimistic about its quality relative to the quality of the incumbents, and their low prescription rates reflected this pessimism.

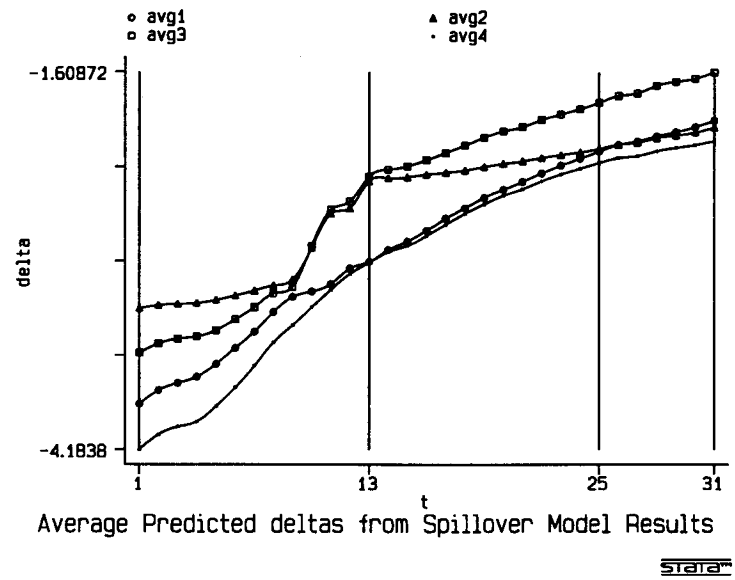
[31]The sole exception is the coefficient on HERFINDAHL PRODUCT LEVEL in the No-Spillovers model.

How much more optimistic do they become on the basis of first-hand experience? This depends on how much doctors revise upward their estimates of the δ's upon prescribing *omeprazole*.

Using the results obtained for the two specifications of the Spillovers model (both with and without controlling for doctor heterogeneity), we have simulated sequences of estimates of the δ's, for each of the 326 doctors in our sample. We averaged these sequences over the doctors, and plotted these average values in figure 3.18. While doctors clearly revise their estimate of *omeprazole*'s quality upward, by the end of the sample period, an average doctor's beliefs about the δ's are still somewhat lower than the "true" values that we have estimated. For example, using the Spillover model results, the average doctor believes that δ_1 is around - 1.8 by the end of the sample period, whereas the "true" value for δ_1 (reading from the results in figure 3.14) is a bit higher, around -0.90.

While the upward trajectories of these graphs are not unexpected, given our estimates, their slopes depend in large part on the amount of first-hand experience — actual prescriptions — that doctors accumulated over the sample period. For example, note that in the upper graph, which uses the estimates for the model without unobserved doctor heterogeneity, the paths of the δ's for diagnosis 2 and 3 cross: near the beginning of the sample period, the average doctor believes that *omeprazole* is most effective for combating pathological hypersecretory conditions (diagnosis 2), but by the end of the sample period, doctors believe on average that *omeprazole* is most effective in attacking peptic ulcers and GERD (diagnosis 3). This ordering contradicts the "true" rankings (discussed above), according to which *omeprazole* is most effective at treating pathological hypersecretory conditions. Part of the reason for this is that, over the sample period, doctors made many more prescriptions for diagnosis 3 then diagnosis 2 — over four times as much, as evidenced in figure 3.5. The paths don't cross for the δ estimates from the model which controls for unobserved doctor heterogeneity, plotted in the lower graph of figure 3.18. How important, though, is first-hand experience versus other non-prescription sources of information, such as advertising?

Answers to this question are given in figures 3.19, 3.20 and 3.21, which decompose (for



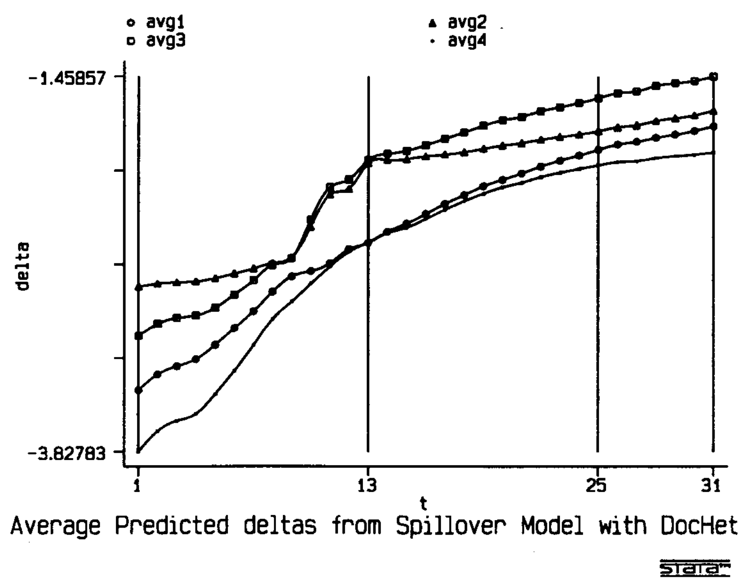


Figure 3.18: Estimated using results from the Spillovers models, and averaged over all 326 doctors in the sample. t is the number of months since entry. avg1=monthly average of $E_t\delta_1$ across doctors; avg2=monthly average of $E_t\delta_2$ across doctors; avg3=monthly average of $E_t\delta_3$ across doctors; avg4=monthly average of $E_t\delta_4$ across doctors; δ_i for $i = 1, \ldots, 4$ denote the "true" qualities of *omeprazole* in treating the different diagnoses.

the 1-Diagnosis, Spillovers models without and with doctor heterogeneity respectively) the across-time changes in the predicted market shares of *omeprazole* into (1) a part due to outside sources of learning (parameterized by advertising and the 31 time dummies in our specifications); and (2) a part due to learning by first-hand experience (i.e., due to the rise in the δ's plotted in figure 3.18). The curve marked avg represents the average doctor in our sample (whose evaluation of the quality of *omeprazole* changes in every period), and the curve marked nopres represents a doctor who never prescribes *omeprazole*, who therefore learns about it only on the basis of advertising and "word of mouth" (proxied by the 31 period-specific dummies) alone. For purposes of comparison, the actual (in-sample) market shares of *omeprazole* for each diagnosis are also plotted, by the curve marked Actual.

It is clear that the actual observed market share path is remarkably well predicted for "average" doctors (the avg lines), but severely underpredicted by doctors who don't prescribe at all (the nopres lines). This implies that the likelihood that a doctor prescribes *omeprazole* does not change much over time unless the doctor actually starts prescribing it, thereby receiving direct verification of its quality relative to the incumbent brands: *first-hand experience plays a* critical *role in overcoming doctors' initial pessimism about the quality of the new entrant.* Keep in mind that these results obtained even though we "stacked the cards" against it by including a full set of time-period dummies to soak up any factors left unexplained by either the learning process or the included covariates: these graphs give very clear evidence that first-hand experience is much more effective than marketing (which here, as in the previous chapter, "substitutes for experience") at overcoming informational barriers to entry into the anti-ulcer drug market: learning through first-hand experience accounts for almost *all* of the rise in *omeprazole*'s in-sample market share over time!

The height of these barriers to entry can be quantified[32] in terms of the extra marketing costs which manufacturers of the entrant *omeprazole*-based drugs had to incur in order to secure a given market share for these drugs. The tru_noad lines give us some idea of the magnitude of these marketing costs. It shows what the path of market shares would have been if the average doctor were perfectly informed about δ, in the absence of advertising

[32] or "pecuniarized", one of my favorite made-up economic terms

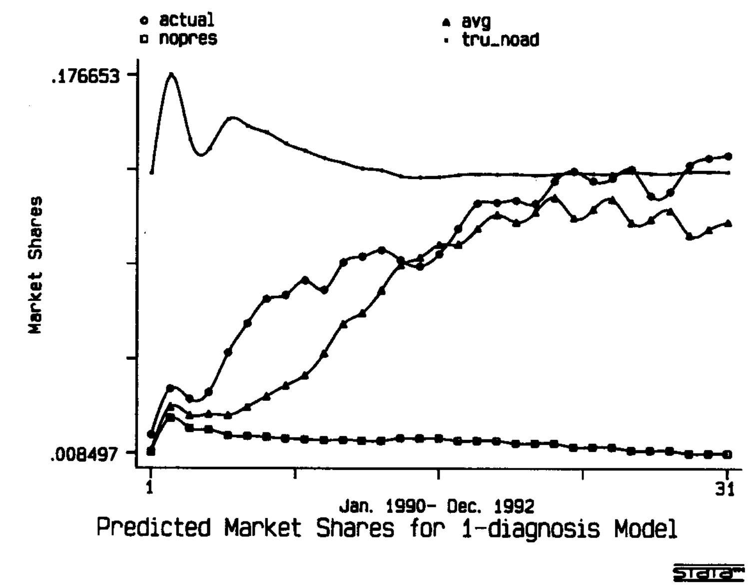


Figure 3.19: **Predicted Market Shares for 1-Diagnosis Model**
On the y-axis, market share of *omeprazole* over all patients over time for 3 scenarios:
(1) **Actual:** actual (in-sample) market share of *omeprazole*
(2) **avg:** predicted market shares for doctor who prescribes the "average" amount of *omeprazole*
(3) **nopres:** predicted market share for doctor who never prescribes *omeprazole* (i.e., who learns about *omeprazole* on the basis of advertising and time dummies alone)
(4) **tru_noad:** predicted market share for doctor in perfect information setting (who knows true δ), but without advertising
Only time dummies, advertising, and δ's vary over time; the other covariates (price and doctor characteristics) are set at their sample means.

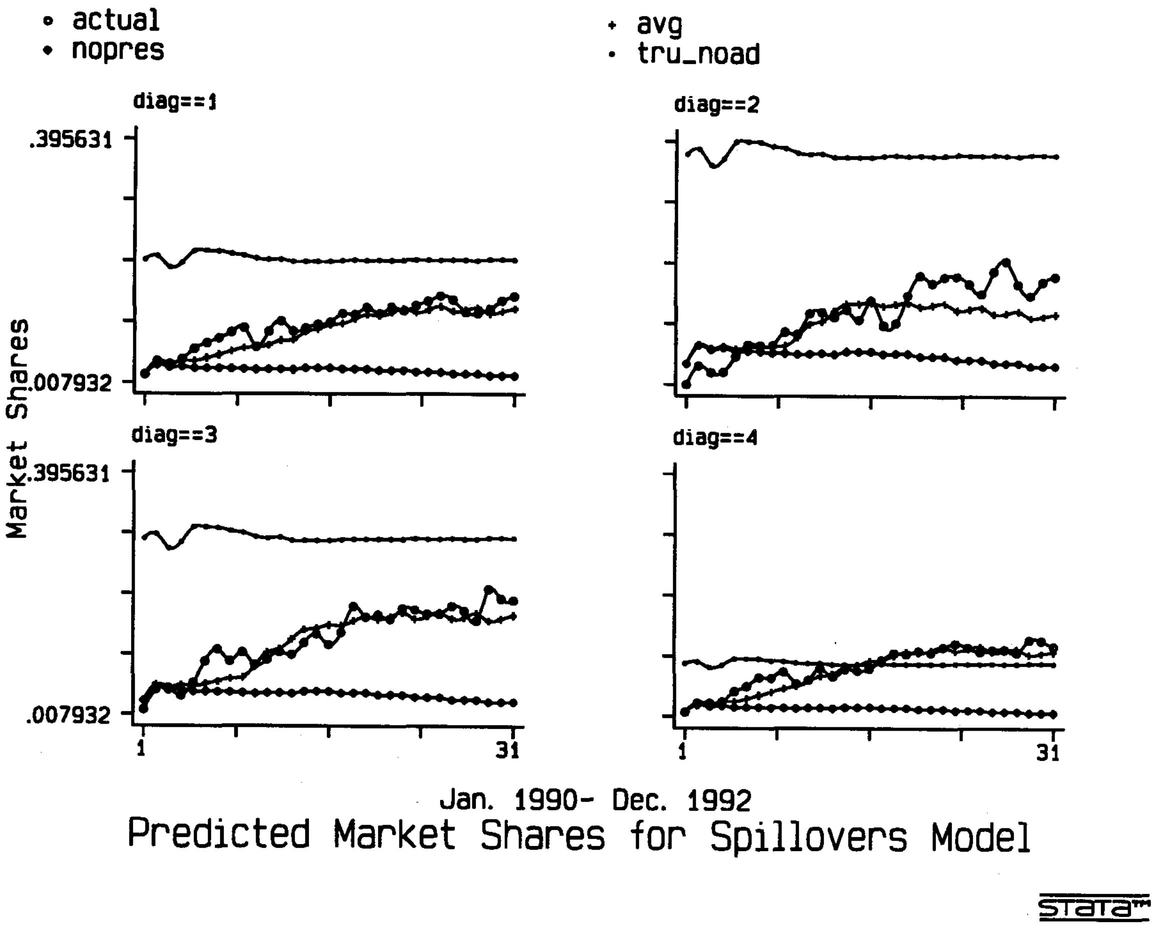


Figure 3.20: **Predicted Market Shares for Spillovers Model without controlling for unobserved doctor heterogeneity**
On the y-axis, market share of *omeprazole* in treating different diagnoses over time for 3 scenarios:
(1) **Actual:** actual (in-sample) market share of *omeprazole*
(2) **avg:** predicted market shares for doctor who prescribes the "average" amount of *omeprazole*
(3) **nopres:** predicted market share for doctor who never prescribes *omeprazole* (i.e., who learns about *omeprazole* on the basis of advertising and the time dummies alone)
(4) **tru_noad:** predicted market share for doctor in perfect information setting (who knows true δ), but without advertising
Only time dummies, advertising, and δ's vary over time; the other covariates (price and doctor characteristics) are set at their sample means.

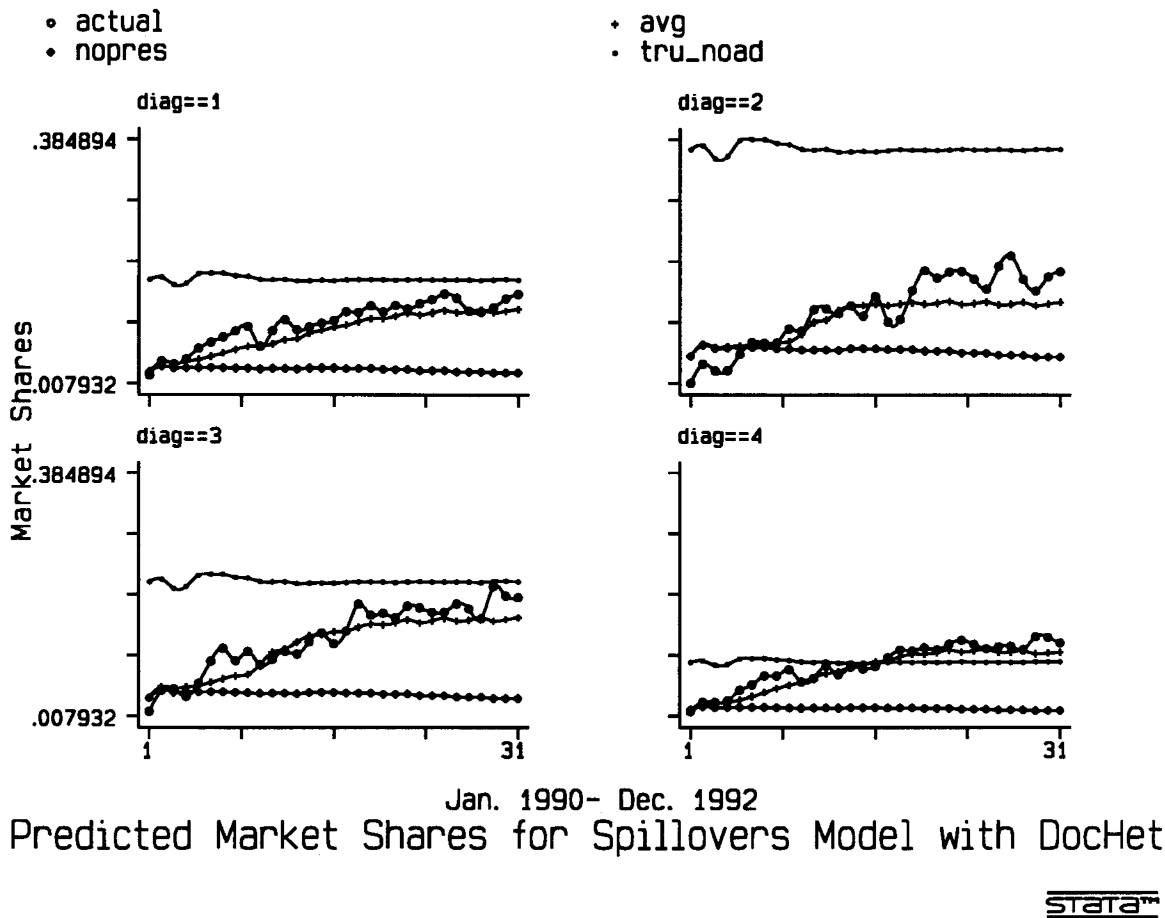


Figure 3.21: **Predicted Market Shares for Spillovers Model after controlling for unobserved doctor heterogeneity**
On the y-axis, market share of *omeprazole* in treating different diagnoses over time for 3 scenarios:
(1) **Actual:** actual (in-sample) market share of *omeprazole*
(2) **avg:** predicted market shares for doctor who prescribes the "average" amount of *omeprazole*
(3) **nopres:** predicted market share for doctor who never prescribes *omeprazole* (i.e., who learns about *omeprazole* on the basis of advertising and the time dummies alone)
(4) **tru_noad:** predicted market share for doctor in perfect information setting (who knows true δ), but without advertising
Only time dummies, advertising, and δ's vary over time; the other covariates (price and doctor characteristics) are set at their sample means.

(here measured in the number of doctors visits made on behalf on a particular drug). For all the diagnoses, the tru_noad line lies above the other lines, which is not surprising, given the finding above that almost all the rise in market share is attributable to the learning process. Clearly, given our estimates, *marketing efforts would be superfluous if doctors were perfectly informed* about the quality differentials of the *omeprazole*-based drugs versus the incumbent drugs.

3.5 Conclusions and Extensions

In this chapter, we have used unique panel data on doctor-level prescription histories to explicitly quantify informational barriers to entry into the anti-ulcer drug market. We focused on the entry of the new molecule *omeprazole* into the Italian market in 1990. Although *omeprazole* represented an innovation upon previous treatments, its market share grew only gradually during our sample period (1990-1992), but by 1996 it became the best selling drug in the world.

We find that *omeprazole*-based drugs suffered a competitive disadvantage upon their entry into the Italian anti-ulcer market due to doctors' initial pessimism about the quality of the drug. Our main result is that even after controlling for time effects, advertising and doctors' heterogeneity, we find that most of the observed gradual growth in demand for *omeprazole* is predominantly explained by doctors' accumulation of positive first-hand experience of these drugs: through prescription of these drug to their patients, doctors become more optimistic about *omeprazole*'s quality, and this increasing optimism drives the gradual rise in *omeprazole*'s market share.

What are the potential welfare effects associated with *omeprazole*'s informational disadvantage relative to a perfect-information world? Several are worth mentioning. First of all, doctors' initial pessimism regarding *omeprazole*'s quality led to far fewer prescriptions in the the early months after its entry than would have been warranted had doctors known *omeprazole*'s "true" quality. Many patients would have been better off had they been switched to *omeprazole* upon entry; this is one source of the welfare loss.

Secondly, as Bagwell [7] points out, the existence of informational barriers to entry may dissuade firms from introducing products into the marketplace which would have been profitable in a perfect information world. This raises the question of whether imperfect information in product markets induces a "suboptimal" (relative to a perfect information setting) selection of products in dynamic market equilibrium. As we pointed out earlier, however, the informational barriers to entry which we have quantified are not "malevolent" in any way — incumbents do not strategically exploit informational advantages to deter entry into markets. These "Bainian" barriers to entry affect all firms equally upon their entry: presumably, the incumbents had to overcome similar informational hurdles when they first entered (or created, as the case may be) the market (BBRU [12] documented how Zantac, the incumbent in our analysis, had to overcome similar hurdles when it entered the market). Therefore, equilibrium market structure is the same as that which would arise in a market characterized by large fixed costs and free entry — sure, fewer products would exist in equilibrium relative to the perfect information (i.e., "zero fixed cost") case, but there is no clear reason that the range of products would be suboptimal in any way.[33] Therefore, informational barriers to entry need not imply dynamic market inefficiencies.

Extensions To conclude, we raise some final caveats and suggest possible extensions. Firstly, in the model presented in this chapter, we assumed that patients within a particular diagnosis class are completely homogeneous. This implies that for a given doctor in a given month, two patients in the same diagnostic class with different prescription histories would have the same chance of being prescribed *omeprazole*. This assumption is restrictive because it doesn't (for instance) allow the noisiness of the signal to decrease (or increase, perhaps, implying *less* informative) for patients who have already tried *omeprazole* often. Therefore, in future work we will focus on incorporating patient-level information —

[33] Free entry in general may lead to a suboptimal *number* of products, as pointed out by Mankiw and Whinston [50]. Furthermore, another way to phrase the argument here is that "history doesn't matter" in the case of informational barriers to entry. This contrasts with the recent literature on network externalities — which lead to entry barriers which are much less severe for the pioneer than for the later entrants — in which path dependence develops and the market can be "locked-in" to an equilibrium characterized by suboptimal products.

especially a patient's prescription history — into the learning model.

Secondly, we do not explicitly model word of mouth among doctors. Standard micro-level diffusion models (see Bass [9], especially pp. 6-7, for a summary, and Lattin and Roberts [48] for a rigorous example) emphasize the importance of local interactions and communication among adopters and non-adopters (or "internal influences") in predicting the aggregate diffusion pattern. We have not modeled these local interactions explicitly here; it is assumed that doctors learn in the privacy of their own offices. Given the requisite data (location of these doctors' practices, at least), developing and estimating such a model would be an interesting extension.

Chapter 4

The Econometrics of English Auctions

In this chapter we derive an econometric model of a multi-round ascending (or English) auction. Much of the previous empirical work on auctions employing structural models (two examples are Paarsch [61] and Laffont, Ossard and Vuong ([46], hereafter LOV)) have focused on *single-round* auctions. Furthermore, the model we develop accommodates *asymmetries* among the bidders in allowing their valuations for an object to be non-identically distributed.

While the aim of this chapter is mostly methodological, we estimate our model using data from the spectrum auctions run by the U.S. Federal Communications Commission (FCC). Our model captures the multi-round aspect of the FCC auctions, but we have abstracted away from the synergies that may result from the simultaneous auctioning off of multiple objects; furthermore, the flexible eligibility rules of these auctions are at odds with the "irrevocable dropout" assumption of the Milgrom-Weber ([57], hereafter MW) model upon which our auction model is based. Therefore, we present the empirical work more to illustrate and suggest solutions to problems which arise in estimating this model in practice than to empirically test economic hypotheses concerning the FCC auctions.

4.1 Single- vs. multiple-round auctions

Consider an auction in which N bidders compete for the possession of a single object. Each bidder has observed a signal concerning the value of the object up for sale; denote these by $(x_1, \ldots, x_N)$. Assume these signals are private, so that bidder i observes only x_i before the auction begins. The actual value of the object to bidder i will be denoted by V_i; this will be a function of $(x_1, \ldots, x_N)$.

The auction model which we focus on in this chapter is an asymmetric version of the "irrevocable dropout" auction described in MW ([57], pg. 1104):

> Initially, all bidders are active at a price of zero. As the auctioneer raises the price, bidders drop out one by one. No bidder who has dropped out can become active again. After any bidder quits, all remaining active bidders know the price at which he quit.

By observing the dropout prices, remaining bidders can subsequently infer the private information possessed by the bidders who have dropped out. This information affects the behavior of the bidders only insofar as the object up for sale has a common value component. In the *independent private values* (IPV) paradigm, for example, where each bidder has a different and private value for the object (which only he knows), the equilibrium bidding strategy is independent of others' valuation: a bidder will bid (up to, in the ascending case) his private valuation of the object.

In *common value* (CV) models, on the other hand, the value of the object up for sale has a component that is the same for all bidders; however, none of the bidders knows the true value of the object, but each observes only a noisy and private signal of the underlying common value. The winner's curse suggests that naive bidders who simply bid their expected valuation of the object conditional only upon their private signal will tend to gain a negative expected profit. In equilibrium, a bidder (call him Mr. Smith) needs to take into account the extra *ex post* information he would gain about losing bidders' private signals which he would gain were he to win the auction. [1]

[1] Loosely speaking, a bidder needs to take into account the consideration that were he to win the auction, he would have observed a signal which is larger in magnitude than the signals of the losing bidders (i.e., he is initially "more optimistic" than the losing bidders).

The main difference between multi-round and single-round auctions lies in the *amount* of information that Mr. Smith would learn about this competitors' private signals were he to win the object, and which he conditions upon in calculating his expected profit from winning — in what follows, we call this *ex post* information. Since the intuition embodied in the equilibrium bidding strategies functions in both the second-price (single-round) and ascending(multi-round) auctions are similar, it is instructive to describe the intuition in the equilibrium of the strategically less complex second-price auction before generalizing to the ascending auction. In this section we mainly develop the intuition of the decision problem faced by the bidders as specified by the equilibrium bid functions for the asymmetric second-price and ascending auctions. For additional details, as well as an equilibrium proof, see Hong [39].

Throughout this chapter, we restrict each bidder i to equilibrium bidding strategies $b_i(x_i)$ which are strictly increasing functions of x_i, bidder i private signal concerning the value of the object.[2] Furthermore, we assume that all the competitors' signals are drawn from distributions which share some region of common support. This rules out extremely asymmetric cases where, say, x_j is drawn from a distribution which support is lower and non-overlapping with the support of x_i. Given the assumption of increasing bidding strategies, bidder j will never win the auction in this case and, indeed, has no incentive to participate in the auction at all![3]

Finally, we assume that the conditional expectations $E[V_i(X_1, \ldots, X_N) \mid X_1, \ldots, X_N]$ (which, in what follows, we will shorten this to $E[V_i \mid X_1, \ldots, X_N]$) is *strictly increasing* in X_i.

[2]This rules out ad-hoc bidding rules such as "stay in no matter what", where $b_i(x_i) = +\infty$ regardless of x_i.

[3]Klemperer [45] discusses several examples of such extremely asymmetric auctions, among them the case of the Los Angeles PCS license, in the FCC auctions, which was handily won by PacBell without much resistance from competitors.

4.1.1 Equilibrium bidding in the second-price auction

Let us consider the decision problem for bidder i in the second-price auction, given his signal x_i, and the equilibrium bid functions for the $N-1$ other bidders $b_j(x_j)$, $j \neq i$, which are common knowledge. Would bidder i regret winning the object at a price of p? If he is risk-neutral, then he doesn't mind winning so long as his *ex post* expected valuation of the object exceeds p. What information does he have *ex post* to form his expectation of the object's value? Naturally, he knows his own private signal x_i, but winning at price p in a second-price auction means that the bidder with the second highest bid has a signal which dictated a bid equal to p, i.e., $\max_{j\neq i} b_j(x_j) = p$. In other words, bidder i is willing to win the object at all prices p for which *ex post* expected revenue is positive:

$$E[V_i \mid x_i, \max_{j\neq i} b_j(x_j) = p] - p \geq 0. \tag{4.1.1}$$

In equilibrium, it can be shown (and we will show it later for the more general ascending auction case) that the assumption of strictly increasing conditional expectations functions guarantees that the expected revenue is strictly decreasing in p. Therefore there is a critical $p^*(x_i)$, a function of x_i, which uniquely satisfies equation (4.1.1) as an equality. This critical $p^*(x_i)$ is a strictly increasing function of x_i, and is in fact bidder i's equilibrium bid $b_i(x_i)$, given x_i:

$$E[V_i \mid x_i, \max_{j\neq i} b_j(x_j) = b_i(x_i)] - b_i(x_i) \;=\; 0.$$

Generalizing, the set of equilibrium bid functions for the N bidders are implicitly defined via the following system of equations:

$$\begin{aligned} b_1(x_1) \quad &= \quad E[V_i \mid x_i, \max_{j\neq 1} b_j(x_j) = b_1(x_1)] \\ \ldots \quad & \quad\quad \ldots \\ b_N(x_N) \quad &= \quad E[V_N \mid x_N, \max_{j\neq N} b_j(x_N) = b_i(x_N)]. \end{aligned} \tag{4.1.2}$$

4.1.2 Equilibrium bidding in the ascending auction

First round equilibrium bid functions Similar intuition motivates the structure of the equilibrium bid functions in the multi-round ascending (or English) auction. The only difference is that in the ascending format, upon winning the auction bidder i knows *all* the private signals of his losing competitors; in other words, his *ex post* information consists of all the signals of his competitors, not just the signal observed by the next-highest bidder, as in the second price auction described above.

Before proceeding, let us introduce the indexing convention we will follow in this paper. N bidders are present in the auction; there will be $N-1$ "rounds" in the auction, indexed $k = 0, \ldots, N-1$. In round 0, all N bidders are active; in round k, only $N-k$ bidders are active. Each round ends when a bidder drops out; bidders are indexed in *decreasing* order of dropout. In other words, bidder N drops out in round 0, and bidder 1 wins the auction; generally, bidder $N-k$ drops out at the end of round k. The dropout prices are indexed by rounds, i.e., $p_0, \ldots, p_{N-1}$. To sum up, bidder i drops out at the end of round $N-i$, at the price p_{N-i}.

Equilibrium bidding strategies in the ascending auction game specify, for each bidder i, bid functions $b_i^k(x_i)$ for each round k, $k = 0, \ldots, N-2$, i.e., $b_i^0(x_i), \ldots, b_i^{N-2}(x_i)$. Given a realization of the private signal x_i, the bid function $b_i^k(x_i)$ tells bidder i which price he should drop out at during round k. We assume that the collections of bid functions $b_i^0(x_i), \ldots, b_i^{N-2}(x_i)$ for bidders $i = 1, \ldots, N$ are common knowledge.

Again, consider bidder i, who is active during round k. As of round k, bidders $N-k+1, \ldots, N$ have already dropped out, at prices $p_k, \ldots, p_0$, respectively. Since the equilibrium bid functions are common knowledge, bidder i can use this information on the identity of the dropout bidders and their dropout prices to infer the private signals $x_{N-k+1}, \ldots, x_N$ observed by these bidders by inverting these bid functions: $x_j = (b_j^{N-j})^{-1}(p_{N-j})$, for $j = N-k+1, \ldots, N$.

Bidder i observes the price level rising; at each price level p he asks himself: would I regret winning the object at price p? As in the second-price auction, he won't regret winning so long as his *ex post* expected valuation of the object exceeds p. In the case of the ascending auction, however, bidder i wins at price p only when *all* the other active bidders in round k (bidders $1, \ldots, N-k$) instantaneously drop out at this price. But if this happens, then bidder i gains the *ex post* information that all the other bidders observed private signals which would have made them drop out at price p. In other words, since the equilibrium strategies are common knowledge, bidder i can infer that the other bidders' signals must be $x_j = (b_j^k)^{-1}(p)$, for $j = 1, \dots, N-k$ and $j \neq i$.

The expected profit to bidder i when he wins the auction at price p during round k is therefore

$$\begin{aligned} E[V_i \mid x_i, x_j = (b_j^k)^{-1}(p), j = 1, \dots, N-k, j \neq i; \\ x_j = (b_j^{N-j})^{-1}(p), j = N-k+1, \dots, N] - p. \end{aligned}$$

If this is positive, then bidder i doesn't mind being active at price p. On the other hand, if this is negative, then bidder i shouldn't be active at price p, during round k. In equilibrium, this expected profit is a decreasing function of the price p.[4] The price p at which he should quit the auction during round k, defined as his *bid function* for round k, is the price $b_i^k(x_i)$ at which he will have a zero expected profit in round k if all other active bidders simultaneously quit at the same price:

$$\begin{aligned} b_i^k(x_i) = \{p :\ E\,[V_i \mid x_i, x_j = (b_j^k)^{-1}(p), j = 1, \dots, N-k, j \neq i; \\ x_j = (b_j^{N-j})^{-1}(p_{N-j}), j = N-k+1, \dots, N\Big] - p = 0.\Big\} \end{aligned} \tag{4.1.6}$$

By monotonicity of the expected profit function in p, there is a unique p, call it p_k^*, which sets expected profit in round k to zero, for each realization x_i. This is illustrated in figure 4.1; the strictly decreasing expected profit function intersects the x-axis at price p_k^*. Therefore

[4] In equilibrium, in round k, bidder i bids a price p such that

$$p = E[V_i \mid (b_i^k)^{-1}(p), (b_j^k)^{-1}(p),\ j = 1, \dots, N-k,\ j \neq i; x_{N-k+1}, \dots, x_N] \tag{4.1.3}$$

(see equation 4.1.6 in main text). On the other hand, if bidder i wins in round k at a price of p, his *ex post* gain is

$$E[V_i \mid x_i, (b_j^k)^{-1}(p),\ j = 1, \dots, N-k,\ j \neq i; x_{N-k+1}, \dots, x_N]. \tag{4.1.4}$$

Given these two equations, we can rewrite expected profit, for a given p, as expression (4.1.4) less the right hand side of equation (4.1.3):

$$\begin{aligned} E\Pi_i^k(p) = E[V_i \mid x_i, (b_j^k)^{-1}(p),\ j = 1, \dots, N-k,\ j \neq i; x_{N-k+1}, \dots, x_N]\ - \\ E[V_i \mid (b_i^k)^{-1}(p), (b_j^k)^{-1}(p),\ j = 1, \dots, N-k,\ j \neq i; x_{N-k+1}, \dots, x_N] \end{aligned} \tag{4.1.5}$$

so that

$$\frac{\partial E\Pi_i^k(p)}{\partial p} = -V_{i1} \cdot \frac{\partial (b_i^k)^{-1}(p)}{\partial p}$$

where V_{i1} denotes the derivative of the conditional expectation with respect to its first argument. By our assumptions above, both V_{i1} and $\frac{\partial (b_i^k)^{-1}(p)}{\partial p}$ are positive, so that $\frac{\partial E\Pi_i^k}{\partial b} < 0$, as claimed.

we can rewrite the above as:

$$
\begin{aligned}
b_i^k(x_i) = E\big[V_i \mid x_i, x_j = b_j^k(x_j) = b_i^k(x_i), j = 1, \ldots, N-k, j \neq i; \\
b_j^{N-j}(x_j) = p_{N-j}, j = N-k+1, \ldots, N\big]
\end{aligned}
\tag{4.1.7}
$$

Generalizing, then, the set of $N-k$ equilibrium bid functions for the bidders $i = 1, \ldots, N-k$

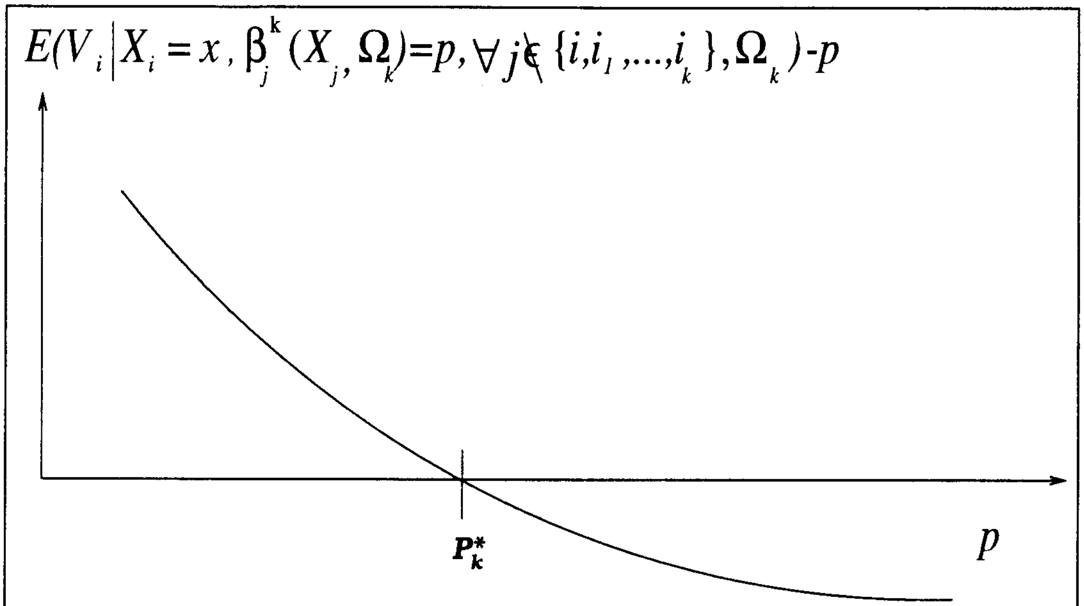


Figure 4.1: Bidder i's decision problem during round k

$\Omega_k \equiv \left\{x_j = (b_j^{N-j})^{-1}(p_{N-j}),\ j = N-k+1, \ldots, N\right\}$

active in round k are implicitly defined via the following system of equations, for any $p > p_{k-1}$:

$$
\begin{aligned}
b_1^k(x_1) &= E\left[V_1 \mid x_1, b_j^k(x_j) = b_1^k(x_1), j = 2, \ldots, N-k, \Omega_k\right] \\
b_2^k(x_2) &= E\left[V_2 \mid x_2, b_j^k(x_j) = b_2^k(x_2), j = 1, 3, \ldots, N-k, \Omega_k\right] \\
&\ldots \\
b_{N-k}^k(x_{N-k}) &= E\left[V_{N-k} \mid x_{N-k}, b_j^k(x_j) = b_{N-k}^k(x_{N-k}), j = 1, \ldots, N-k-1, \Omega_k\right]
\end{aligned}
\tag{4.1.8}
$$

where $\Omega_k \equiv \left\{x_j = (b_j^{N-j})^{-1}(p_{N-j}), \text{ for } j = N-k+1, \ldots, N\right\}$. The full set of equilibrium bid functions is analogously described by sets of $N-k$ equations for each round $k = 0, \ldots, N-1$.[5]

[5]From a decision-theoretic point of view for each individual bidder, the structure of the equilibrium bid

In Wilson [73], this conditioning event $b_j^k(x_j) = b_i^k(x_i)$ in bidder i equilibrium bid function for round k is described as bidder i's equilibrium conjecture that all other active bidders j have the same targeted dropout price as he (bidder i) does. The discussion so far has shown that this equality arises because if bidder i wins the object at price $b_i^k(x_i)$, it must be that all other remaining bidders j have dropped out at this price, allowing bidder i to infer *ex post* that all these bidders had private signals which caused them to drop out at this price, i.e., $x_j = (b_j^k)^{-1}(b_i^k(x_i)) \leftrightarrow b_j^k(x_j) = b_i^k(x_i)$. Given the bidder asymmetries, this is *far* from an assertion that in equilibrium, bidder i believes that his remaining competitors have the same private signal as he does.[6]

From a computational point of view, the structure of these equilibrium bid functions is particularly attractive since the conditioning events are *points* rather than *sets*. In the case of the latter, the conditional expectations in (4.1.8) would involve multi-dimensional integrals, which can be difficult to compute.

4.2 Log normal asymmetric ascending auction model

An econometric ascending auction model would use observed dropout prices to recover the parameters of the equilibrium bid function, which in turn depend on the preferences for the objects being sold, which vary across bidders. Formulating such a model involves first and foremost specifying distributions on the latent signals and valuations $X_1, ..., X_N, V_1, \dots, V_N$ such that the resulting mapping (i.e., the equilibrium bid function) from observed dropout prices to unobserved valuations is tractable enough to apply data to and estimate.

Difficulties arise in doing this because the updating process introduces a massive amount

function shares some interesting similarities with the optimal policy in an *optimal stopping* problem, in which the "state variable" is the conditional expectation of the object's value (EV_i) and time is measured by the (deterministically rising) price level.

Note, however, that the "law of motion" for the state variable EV_i is not stationary, i.e., functionally independent of our time measure P: how EV_i evolves as P rises depends on the number (and identities) of the bidders who have already dropped out before price P, as well as the number and identities of the remaining bidders. Because of this, the optimal "stopping policy" would also be non-stationary, in the sense that the "critical value" of the state under which bidder i should drop out — this is his bid function — is not constant over the course of the auction. gain extra information over the course of the auction —

[6]This is true only for the symmetric case, where $b_j^k(x_j) = b^k(x)$, $\forall\, j = 1, \dots, N$. The conditioning event $b_j^k(x_j) = b_i^k(x_i)$ then simplifies to $x_j = x_i$.

of recursivity into the definition of the bid function. For example, assume four bidders (A,B,C,D) and assume the first three drop out in rounds 1,2, and 3, respectively. After bidder A drops out, the remaining bidders (B,C,D) invert the equilibrium bid function for bidder A, in order to obtain his private signal x_A. In round three, bidder C and D must invert bidder B's bid function during round two, which is her expected value of the object in round two, *conditional not only on her private signal but also on x_A which she inferred by inverting bidder A's conditional expectation function from the previous round.*

The recursive structure which results (involving conditional expectation — in computational terms, integrals — functions which have as arguments inversions of other conditional expectation functions which are themselves inversions of other conditional expectation functions) quickly becomes intractable if the conditional expectations derived during the updating process become too complicated. Thus the challenge in implementing this model econometrically lies in picking a conjugate family of distributions for the latent variables (V, X) such that the resulting conditional expectation functions have tractable functional forms which facilitate inversion.

Fortunately, if we assume that bidders' valuations are log-normally distributed, we can derive closed-form formulas for the conditional expectation functions. This section contains the main steps in the derivation of these dropout prices given this assumption. Suppose N bidders are contending for an single object in an ascending auction. The value of the object to bidder i is assumed to be $V_i = A_i \times V$. A_i is a bidder-specific private value for bidder i. V is the common market value which is not known to any of the bidders. In other word, V_i is the product of a common value part and a private value part.

The V and the A_i's are independently log normally distributed. If we let $v = \ln V$ denote the natural log of V, and similarly let $a_i = \ln A_i$. Then

$$v = m + \epsilon_v \sim N(m,\ r_0^2)$$

$$a_i = \bar{a}_i + \epsilon_{a_i} \sim N(\bar{a}_i,\ t_i^2).$$

Each bidder is assumed to have a noisy signal of the value of the object, X_i, which has the form $X_i = A_i \times E_i$. Here E_i is a noisy estimate of the common market value V.

$E_i = V \times \exp\{s_i\xi_i\}$ in which ξ_i is an (unobserved) error term that has a normal distribution with mean 0 and variance 1. If we let $v_i = \ln V_i$[7] and $x_i = \ln X_i$, then conditional on v_i, $x_i = v_i + \epsilon_{e_i} \longrightarrow N(v_i,\ s_i^2)$. Note that bidder i "observes" both A_i and E_i, but if he drops out, all that the other bidders "observe" is X_i, the product of the two.

We can summarize the information structure by the following equations:

$$\begin{array}{lllll} v & & & = & m + \epsilon_v \\ v_i & = & a_i + v & = & \bar{a}_i + m + \epsilon_v + \epsilon_{a_i} \\ x_i & = & v_i + \epsilon_{e_i} & = & \bar{a}_i + m + \epsilon_v + \epsilon_{a_i} + \epsilon_{e_i} \\ v_j & = & a_j + v & = & \bar{a}_j + m + \epsilon_v + \epsilon_{a_j} \\ x_j & = & v_j + \epsilon_{e_j} & = & \bar{a}_j + m + \epsilon_v + \epsilon_{a_j} + \epsilon_{e_j} \end{array} \tag{4.2.9}$$

All the ϵ's are normally distributed with zero mean. The variance for the ϵ_{a_i}'s are denoted t_i. The variance for the ϵ_{e_i}'s are denoted s_i. Let $r_i = \sqrt{t_i + s_i}$. Denote the variance for ϵ_v by r_0.

The joint distribution of $(V_i, X_i, i = 1, \ldots, N) = \exp(v_i, x_i, i = 1, \ldots, N)$ is fully characterized by $\{m, \bar{a}, t, s, r\}$ where $\bar{a}$ is the collection of $\bar{a}_i$'s, similarly for t, s, and r. These are all common knowledge among the bidders.

4.2.1 Deriving the equilibrium bid functions

Before describing the process whereby we derive the equilibrium bid function, note that given the log-normality assumptions on V_i and the X_i's, $i = 1, \ldots, N$, the conditional expectation functions for V_i take the following form:

$$E[V_i \mid X_1, \ldots, X_N] = \exp\left(E(v_i \mid x_1, \ldots, x_N) + \frac{1}{2}Var(v_i \mid x_1, \ldots, x_N)\right), \tag{4.2.10}$$

for $i = 1, \ldots, n$. Given that $(v_i, x_1, \ldots, x_N)$ are jointly normal, the conditional mean function $E(v_i \mid x_1, \ldots, x_N)$ is linear in $x_1, \ldots, x_N$ and $Var_k(v_i \mid x_1, \ldots, x_N)$, the variance of v_i conditional on $x_1, \ldots, x_N$, is constant and does not depend on $x_1, \ldots, x_N$. This follows from the familiar conditional mean and variance formulas for jointly normal random variables[8]. Therefore the conditional expectation functions for $V_i \equiv \exp(v_i)$ are log-linear.

[7] Recall from above that V_i is the true value of the object for bidder i.

[8] They are reproduced in equations (3.2.14) and (3.2.15) in chapter 3.

In round k of the auction, k bidders have dropped out. Recall that in equilibrium (as discussed above), the system of $(N-k)$ bid functions (for the $n-k$ remaining bidders) are conditional expectation functions (reproduced from equations (4.1.8):

$$
\begin{aligned}
b_1^k(x) &= E[V_1 \mid X_1 = x, b_2^k(X_2) = b_1^k(X_1), \ldots, b_{N-k}^k(X_{n-k}) = b_1^k(X_1), \Omega_k] \\
b_2^k(x) &= E[V_2 \mid b_1^k(X_1) = b_2^k(X_2), X_2 = x, \ldots, b_{N-k}^k(X_{n-k}) = b_2^k(X_2), \Omega_k] \\
&\ldots \\
b_{N-k}^k(x) &= E[V_{N-k} \mid b_1^k(X_1) = b_{N-k}^k(x), b_2^k(X_2) = b_{N-k}^k(x), \ldots, X_{N-k} = x, \Omega_k].
\end{aligned}
$$

In period k, bidders $N-k+1, \ldots, N$ have dropped out, revealing $X_{N-k+1}, \ldots, x_N$. The equilibrium conjectures stated above mean that, essentially, bidder i conjectures that $x_j = (b_j^k)^{-1}(b_i^k(x_i))$, $\forall\ j = 1, \ldots, N-k,\ j \neq i$. $(b_j^k)^{-1}(P)$ indicates the inverse of $b_j^k(\cdot)$ evaluated at an arbitrary price P, i.e., what bidder j's type $(b_j^k)^{-1}(P)$ would be if he chooses to drop out at price P.

Given this notation, we can rewrite the conditional expectations above as functions of the inverse bid functions $((b_1^k)^{-1}(P) \ldots (b_{n-k}^k)^{-1}(P))$, for an arbitrary P, rather than functions of $(x_1 \ldots x_{N-k})$:

$$
\begin{aligned}
P &= E[V_1 \mid X_1 = x, (b_2^k)^{-1}(P), \ldots, (b_{N-k}^k)^{-1}(P), \Omega_k] \\
P &= E[V_2 \mid (b_1^k)^{-1}(P), X_2 = x, \ldots, (b_{N-k}^k)^{-1}(P), \Omega_k] \\
&\ldots \\
P &= E[V_{N-k} \mid (b_1^k)^{-1}(P), (b_2^k)^{-1}(P), \ldots, X_{N-k} = x, \Omega_k].
\end{aligned}
\tag{4.2.11}
$$

Given a value for P, (4.2.11) defines a set of $(N-k)$ nonlinear equations with $(b_1^k)^{-1}(P), (b_2^k)^{-1}(P), \ldots, (b_{N-k}^k)^{-1}(P)$ as the $(N-k)$ "unknowns". In this form, then, we can solve for the set of $(N-k)$ inverse bid functions for period k from that period's system of conditional expectation functions pointwise for each P, given Ω_k in the range of the bid.

Using the conditional mean formula for log-normal variates (4.2.10) and the conditional mean and variance formulas for jointly normal variates, the formulas in (4.2.11) take the following form (where $p = \log P$):

$$\begin{aligned} p =& \frac{s_i^2/r_i^2 + t_i^2 h}{r_i^2 h} x_i + \frac{s_i^2}{r_i^2 h} \sum_{j=1, j \neq i}^{N-k} \frac{(b_j^k)^{-1}(p) - a_j + a_i}{r_j^2} \\ &+ \frac{s_i^2}{r_i^2 h} \left(\frac{m + a_i}{r_0^2} + \sum_{l=N-k+1}^{N} \frac{x_l - a_l + a_i}{r_l^2} \right) + \frac{1}{2} \frac{r_i^2 s_i^2 t_i^2 h + s_i^4}{r_i^4 h} \end{aligned} \tag{4.2.12}$$

for each remaining bidder $i = 1, \dots, N - k$.

If we invert this $(N - k)$ system of equations, we can solve for $(b_j^k)^{-1}(p)$ as a function of p and $x_{k+1} \dots x_N$, for $j = 1 \dots N - k$, in each round k of the auction. In fact, these equations are the log inverse bid functions $(b_j^k)^{-1}(p)$ for each of the remaining bidders:

$$\left((b_l^k)^{-1}(p), \dots, (b_{N-k}^k)^{-1}(p)\right)' =$$

$$\begin{pmatrix} \frac{s_1^2/r_1^2+t_1^2 h}{r_1^2 h} & \frac{s_1^2}{r_1^2 r_2^2 h} & \cdots & \frac{s_1^2}{r_1^2 r_{n-k}^2 h} \\ \frac{s_2^2}{r_1^2 r_2^2 h} & \frac{s_2^2/r_2^2+t_2^2 h}{r_2^2 h} & \cdots & \frac{s_2^2}{r_2^2 r_{n-k}^2 h} \\ \vdots & \vdots & \ddots & \\ \frac{s_{N-k}^2}{r_1^2 r_{N-k}^2 h} & \frac{s_{N-k}^2}{r_2^2 r_{N-k}^2 h} & \cdots & \frac{s_{N-k}^2/r_{N-k}^2+t_{N-k}^2 h}{r_{N-k}^2 h} \end{pmatrix}^{-1} \begin{bmatrix} p \\ p \\ \vdots \\ p \end{bmatrix} + \begin{bmatrix} f_1(x_{N-k+1}, \dots, x_N) \\ f_2(x_{N-k+1}, \dots, x_N) \\ \vdots \\ f_{N-k}(x_{N-k+1}, \dots, x_N) \end{bmatrix} \tag{4.2.13}$$

where $f_1, f_2, \dots, f_{N-k}$ are all linear functions of $(x_{N-k+1}, \dots, x_N)$ which do not depend on p at all. Hence all the $((b_1^k)^{-1}(p), \dots, (b_{N-k}^k)^{-1}(p))$ are linear functions of p. After deriving the inverse bid functions in this manner, we can easily invert them to get the bid functions.

In order for equilibrium to exist (and for the inverse bid functions to be well-defined), we require that the bid function are increasing in x. For the log-normal case considered in this chapter, it turns out that we can directly verify (from analytically inverting the matrix in the above system of equations) that the inverse bid functions are increasing in p. This in turns implies that the bid functions themselves are increasing in x; therefore, the monotonicity conditions required for existence of equilibrium are satisfied.

4.2.2 Deriving the joint density of dropout price vector

The likelihood function for a given auction is the (multivariate) density of the observed set of dropout prices from that auction. Given distributional assumptions on the x's and the nature of equilibrium in the ascending auction game, the "regression equation" which we want for a given dropout price is

$$P_k = f(x_{N-k}; x_{N-k+1} \dots x_N); \tag{4.2.14}$$

in other words, the dropout price in round k as a function of bidder $N-k$'s valuation x_{N-k},[9] as well as all the information he has in round k, which are the valuations $(x_{N-k+1} \dots x_N)$ of the bidders who have dropped out before round k. This "regression equation" specifies the transformation from the unobserved latent variables (the x's) which we have made distributional assumptions about to the observed variables (the dropout prices).

However, the system of equations in 4.2.12 give p_i as a function of *all* of the x's. Therefore, in order to derive the regression equations in 4.2.14, we must go through two steps: first, we solve for x_k as a function of $(p, x_{k+1}, \dots, x_N)$, where p is a generic p, for rounds $k = 1, \dots, N-1$[10]; secondly, we invert these $(N-1)$ functions to get p_k as a function of $(x_k, \dots, x_N)$, as desired. We go into each step in more detail.

Step 1 In the first step, we loop over each round k of the auction in order to solve for x_k, bidder k's valuation, as a function of p_k and $x_{k+1} \dots x_N$, the valuations of the bidders who have already dropped out (here we continue the indexing convention that bidder N drops out first). This is done using the methodology described in the previous section.

In round 0, for example, none of the x's are known (since no one has dropped out yet) so using the entire system of equations in 4.2.12 (N equations in N unknowns) we can solve for $(b_j^1)^{-1}(p)$ as a function of just a generic p, for $j = 1 \dots N$. From these equations, we keep only the one for bidder N — $(b_n^1)^{-1}(p)$ — the bidder who drops out during round 0. This equation is the inverse bid function for bidder N during the round that he drops out. We keep only one equation from each round because we only observe dropout prices (more fancifully, the econometrician only "observes" a bidder's bid function when he drops out).

Similarly, in round k, we solve for $(b_j^k)^{-1}(p)$ as a function of generic p and $x_{k+1} \dots x_N$, for $j = 1 \dots N-k$, and keep the equation for bidder $n-k$. So after looping through all

[9]Recall that by our indexing convention, bidder $N-k$ drops out in round k, at the price of p_k.
[10]since we don't observe the "dropout price" for the winning bidder

rounds $k = 0 \ldots N-2$, we are left with the $(N-1)$ equations: namely, x_k as a function of p and $x_{k+1} \ldots x_N$, for $k = 0, \ldots, N-2$.

Step 2 Secondly, we invert each of these equations singly in order to get p_k, the dropout price during round k, as a function of $x_k \ldots x_N$, for $k = 0 \ldots N-2$. Given our assumptions about the means and variances of the x's, we can analytically derive the vector of means $\vec{\mu}_p(\theta, z)$ and the covariance matrix $\Sigma_p(\theta, z)$ for these $N-1$ equations. θ is shorthand for the parameter vector and z are the included covariates (both described below). The likelihood function for the observed vector of dropout prices $\mathcal{P}$[11] for a given auction is then multivariate normal, with mean $\vec{\mu}_p$ and covariance Σ_p:

$$f(\mathcal{P}, \theta, z) = (2\pi)^{-\frac{n}{2}} |\Sigma_p(\theta, z)|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathcal{P} - \mu_p(\theta, z))' \Sigma_p(\theta, z)^{-1} (\mathcal{P} - \mu_p(\theta, z))\right] \quad (4.2.15)$$

More detail on the likelihood function derivation, as well as formulae for $\vec{\mu}_p$ and Σ_p, is given in the appendix to this chapter.

4.2.3 Equilibrium truncation conditions

The multivariate normal density in equation 4.2.15 does not fully impose all the equilibrium restrictions implied in our ascending auction model. More specifically, the joint density of the dropout prices in equation 4.2.15 conditions on the *observed dropout order*, i.e., the first element of $\mu_p(\theta, z)$ is derived under the assumption that bidder 1 is the observed winning bidder, the second element is derived under the assumption that bidder 2 is the observed second-highest bidder, etc. However, there are regions in the multivariate normal density in (4.2.15) from which vectors of prices could be drawn which imply a dropout order different than the one which is observed (e.g., bidder $N-k$ drops out before round k), which would be inconsistent with the structure of the multivariate normal density we have derived (e.g., . the $N-k$th element of $\mu_p(\theta, z)$ is a function of bidder $N-k$'s covariates).

Therefore, the nature of equilibrium in this game imposes a set of truncation conditions on the joint distribution of the observed sequence of dropout prices in a given auction these conditions state that, in each round of the auction, given the parameters θ, the targeted

[11] which doesn't include the extra piece of information that the winner's targeted dropout price in the last round is greater than the observed winning price.

dropout prices of the remaining bidders for that round must be higher than the dropout price at that round. This ensures, roughly speaking, that bidder $N-k$ will not drop out before round k.

To be more specific, the multivariate normal density (4.2.15) only imposes the condition that a bidder who drops out during round k, at a price $p_k(\theta, z)$, has an expected valuation of the object equal to $p_k(\theta, z)$:

$$E[v_{N-k} \mid x_{N-k}, (b_1^k)^{-1}(b_{N-k}^k(x_{N-k})), \dots, (b_{N-k-1}^k)^{-1}(b_{N-k}^k(x_{N-k}));$$
$$x_{N-k+1}, \dots, x_N] =$$
$$p_k(\theta, z) = b_{N-k}^k(x_{N-k}; \theta, z).$$

In what follows, we suppress z as an argument in $p_k(\cdot)$ and $b_j^k(\cdot)$.

To ensure that the "correct" dropout order occurs, we need to impose that, at the given parameter values, all remaining bidders $j = 1, \dots, N-k-1$ have expected valuations greater than $p_k(\theta)$:

$$b_j^k(\theta) = E[v_j \mid x_j, (b_l^k)^{-1}(b_j^k(x_j)),\ l = 1, \dots, N-k-1,\ l \neq j;\ x_{N-k+1}, \dots, x_N] > p_k(\theta),$$

for all $j = 1, \dots, N-k-1$, the bidders who are left in the auction after round j.

In other words, $b_{N-k}^k(x_{N-k}; \theta) < b_j^k(x_j; \theta)$, for $j = 1, \dots, N-k-1,\ k = 0, \dots, N-2$, where $b_j^k(x_j; \theta)$ denotes the targeted dropout price of bidder j in round k, at the current values of the parameters θ. These constraints can only be stated in terms of the model parameters; therefore they violate the standard regularity condition required for consistency and asymptotic normality of the maximum likelihood estimator.

The truncation probability is then equal to $P(\mathcal{A})$, where

$$P(\mathcal{A}) = \text{Prob(the truncation conditions hold)}.$$

Then the likelihood function is the sum of $NOBS$ terms, where $NOBS$ represents the number of auctions conducted. Each of the $NOBS$ terms is the multivariate normal density of the observed dropout prices divided by the truncation probability:

$$L(\mathcal{P} \mid \theta, z) = \begin{cases} \dfrac{f(\mathcal{P}, \theta, z)}{\int_{\mathcal{A}} f(\mathcal{P}, \theta, z)\, d\mathcal{P}} & \text{if truncation conditions hold.} \\ 0 & \text{otherwise.} \end{cases} \tag{4.2.16}$$

where θ denotes the parameter vector and z the covariates.

4.3 Estimation issues

4.3.1 Simulating the truncation probability

Using the density of the dropout prices that we have derived above (in equation 4.2.16) as the basis for estimation poses various problem. From a computational point of view, the integral in the denominator, which indicates the truncation probability, integrates the multivariate normal density over several dimensions and cannot be analytically evaluated. Numerical integration is also difficult. Therefore, in both the maximum and nonlinear least squares estimation methodologies considered below, we simulate the truncation probabilities.

For each fixed value of the parameter vector, we draw a normal random vector of the bidders' private valuations (the x's) for each auction, using the estimated mean and variance-covariance of the bidders' private valuations, and calculate all the targeted dropout prices at all rounds. Then we check all of the truncation inequalities. If all of them are satisfied, we set an indicator to be 1. If one of these conditions fails, we set an indicator to be zero. We repeat this random drawing for the number of simulations for each auction. The simulated truncation probability, is then the average of the indicators over the number of simulations.

The problem with the crude simulator is that the simulated probability is not smooth in the parameters. This can create severe problems for the optimization routines. To reduce non-smoothness we use an independent probit kernel smoother to smooth the indicator functions (see McFadden [54] for more details). For each auction t, and each simulated draw, we calculate the smoothed indicator function as:

$$I_t = \Pi_{k=0}^{N_t-2} \Pi_{j=1}^{N_t-k-1} \Phi\left(\frac{p_j^k - p_{N_t-k}^k}{h}\right)$$

where p_j^i is the targeted dropout price for bidder j in round i, and h is a small number chosen to be the bandwidth of the smoothing.

4.3.2 Maximum likelihood

The first estimation alternative which we consider is maximum likelihood estimation, using the derived density of the log dropout prices in (4.2.16) as our likelihood function. However, several problems are inherent in this approach, which motivate the nonlinear least squares methodology described further below.

Non-regular maximum likelihood problem The equilibrium truncation conditions described above involve model parameters. Since this implies that the support of the observables (the dropout prices) depends on model parameters,[12] the resulting maximum likelihood estimation problem is nonregular. We are not completely clear about the consistency and asymptotic distribution of the estimates obtained by maximizing the likelihood function (4.2.16).[13] Therefore the standard errors reported below[14] should be considered more as indicators of the "stability" of the likelihood function in the neighborhoods of the estimated parameters rather than used as the basis for rigorous statistical inference.

Penalized likelihood approach If we set the value of the likelihood function to zero when the truncation conditions are violated, the objective function becomes very nonlinear, and estimation becomes difficult from a computational point of view. Rather than doing this, therefore, during each iteration of the likelihood function we calculate the targeted dropout prices for each of the non-dropout bidders during each round of the auction at the current parameter values and calculate the extent that they violate the truncation conditions. In other words, when the truncation conditions are satisfied, we don't do anything to the log likelihood functions; whenever one of the truncation conditions is violated, we subtract from the log likelihood function a large factor multiplied by the square of the deviation. For example, during round k, at the current parameter values, if the targeted dropout price for bidder $N-k$ (who drops out in that round) is P_{n-k}^k, and the targeted dropout price for bidder $N-k-1$, who remains after that round, is p_{N-k-1}^k, then we check if $p_{N-k}^k \leq p_{N-k-1}^k$ holds. If this holds, then we don't "penalize" the log likelihood function.

[12] Usual truncated regression models assume that the truncation points are known constants. For example, the Tobit labor supply model assumes that an agent works only when his wage exceeds some reservation wage, normalized to be zero. This implies that the observed wages are truncated at zero, a known constant.

[13] A fuller discussion is given in Hong [40], where consistency and asymptotic distributions have been derived for some other nonregular maximum likelihood problems, including the symmetric first-price auction model. For multivariate dependent variables, such as the vector of dropout prices observed for a given ascending auction, there is no general theory available to guide estimation.

[14] The covariance matrix of the estimates was calculated using standard "outer-product of the gradient" of the likelihood function evaluated at the estimated parameter values.

But if this fails, then we subtract

$$\alpha \times \left(p^k_{N-k} - p^k_{N-k-1}\right)^2$$

from the log likelihood function, where α is some large positive number. Equivalently, we can subtract

$$\alpha \times \max\left(p^k_{N-k} - p^k_{N-k-1}, 0\right)^2$$

from the log likelihood function. By this penalization we are seeking to push the parameter estimates to the regions for which the truncation constraints are satisfied. After convergence of this penalized likelihood estimation procedure, we count the number of truncation conditions violated by the converged parameter estimates. When this number is zero, we effectively obtained the constraint maximum. If this number is not zero, we can always increase the penalizing factor to push it to zero.

To recover the targeted dropout prices for remaining bidders for each round of the auction, we first need to recover the bidders' signals at the current parameter values. These are recovered recursively. The first dropout bidder's dropout price is a function of model parameters and his private signal. His private signal is then a function of his dropout price and the parameters. The second dropout bidder's dropout price is a function of parameters, his private signal, and the first dropout-bidder's revealed private signal. Then after solving for the first bidder's private signal, we could solve for the second dropout bidder's private signal as a function of the second dropout bidder's dropout price, model parameters, and the first bidder's revealed private signal. Iterating this procedure we can recover all bidders' private signals. Then from these bidders' private signals and the parameter estimates we can calculate all the targeted dropout prices at each round of the auction.

Thus our final penalized likelihood function takes the following form:

$$logL_1 + \alpha \sum_{t=1}^{NOBS} \sum_{k=1}^{N_t-2} \sum_{j=1}^{N_t-k-1} \left[\max\left(b^k_{N-k}(x_{N-k};\theta) - b^k_j(x_j,\theta)\right), 0\right]^2 \tag{4.3.17}$$

where $logL_1$ is log of the likelihood function in (4.2.16).

4.3.3 Simulated Nonlinear Least Squares Estimation

An alternative which avoids the nonregularity problem in estimating the model via maximum likelihood is simulated nonlinear least squares (SNLS) estimation, based on the methodology of LOV [46]. This estimator minimizes the usual nonlinear least squares (NLS) objective function

$$Q_T(\theta) = (1/T) \sum_{t=1}^{T} \sum_{k=0}^{N_t-2} \left(p_k^t - m_k^t(\theta)\right)^2 \tag{4.3.18}$$

where p_k^t is the kth observed log dropout price for auction t, and $m_k^t(\theta)$ is its corresponding expectation (or first moment, therefore denoted as m), taken with respect to its density as given in equation 4.2.16. Because $m_k^t(\theta)$, the mean of a multivariate truncated normal distribution, is difficult to compute analytically, we replace $m_k^t(\theta)$ in equation 4.3.18 by an unbiased simulation estimator $\tilde{m}_k^t(\theta) = \frac{1}{S}\sum_{s=1}^{S} p_{k,s}^t(\theta)$, where S is the number of simulation draws and $p_{k,s}^t$ is the simulated dropout price using the sth draw. The ensuing Simulated Nonlinear Least Squares (SNLS) objective function

$$Q_{S,T}(\theta) = \frac{1}{T} \sum_{t=1}^{T} \sum_{k=0}^{N_t-2} \left(p_k^t - \tilde{m}_k^t(\theta)\right)^2 \tag{4.3.19}$$

yields a consistent estimate of θ when $S \longrightarrow \infty$.[15]

Simulating $\mathbf{m}_k^t(\theta)$ To be more specific, the first moment $m_k^t(\theta)$ of the kth dropout price p_k^t, for $k = 0, \ldots, N_t - 2$, is

$$m_k^t(\theta) = \int_{\vec{x}} p_k^t(\vec{x};\theta)\, \mathbf{1}(\vec{x} \in A(\theta))\, \frac{f_t(\vec{x};\theta)}{Pr(A(\theta))} d\vec{x}. \tag{4.3.20}$$

where $\vec{x} \equiv \{x_1, \ldots, x_{N_t}\}$ denotes the vector of the signals of bidders in the order of dropping out, $p_k^t(\vec{x};\theta)$ specifies the kth dropout price as a function of the parameters and realized vector of bidder signals. $A(\theta)$ denotes the event that the observed order of dropping out is realized. The integration is over the N_t-dimensional vector of bidder signals.

[15] For a fixed value of S, the probability limit of the SNLS objective function $Q_{S,T}$ is not equal to Q_T, the NLS objective function. Therefore the estimates of θ will be inconsistent. Unfortunately, in our case, we cannot introduce a bias-correction term into the SNLS objective function, as in LOV ([46], pg. 959), because the corresponding bias-correction term in our case (i.e., the function $\Delta_{S,T}(\theta)$ which sets $\mathrm{plim}_{T\to+\infty}(Q_{S,T}(\theta) + \Delta_{S,T}(\theta) - Q_T(\theta)) = 0$ for fixed S) is difficult to calculate.

Difficulties arise in minimizing an objective function which is calculated from random variables drawn directly from a truncated distribution where the truncation conditions (equivalently, the support) depend of the parameter values θ, as in our case. The SNLS objective function will not be very smooth in the parameters θ since, Using standard "acceptance-rejection" methods to sample from a truncated distribution, small changes in the estimates of θ can lead some formerly "rejected" draws to be "accepted".

In overcoming this problem, we employ *importance sampling.* Instead of drawing random variables directly from the multivariate truncated normal distribution of signals which depends on model parameters, we make random draws of $\vec{x}$ from a multivariate density function $g(\vec{x};\theta)$ which, while similar in shape to the truncated multivariate normal distribution which we wish to sample from, does not have a support which depends on the parameter values. We can transform the expression in (4.3.20):

$$m_k^t(\theta) = \int_{\vec{x}} p_k^t(\vec{x};\theta)\,\mathbf{1}(\vec{x}\in A_t(\theta))\,\frac{f_t(\vec{x};\theta)}{g(\vec{x})\,Pr(A_t(\theta))}g(\vec{x})\,d\vec{x}. \tag{4.3.21}$$

Furthermore, $Pr(A(\theta))$ is also difficult to evaluate directly but can also be simulated by the same random draws:

$$Pr(A_t(\theta)) = \int_{\vec{x}} f_t(\vec{x};\theta)\,\mathbf{1}(\vec{x}\in A_t(\theta))\,d\vec{x} = \int_{\vec{x}} \frac{f_t(\vec{x};\theta)\,\mathbf{1}(\vec{x}\in A_t(\theta))}{g(\vec{x})}g(\vec{x})\,d\vec{x}$$

Therefore,

$$m_k^t(\theta) = \left[\int_{\vec{x}} p_k^t(\vec{x};\theta)\,\mathbf{1}(\vec{x}\in A_t(\theta))\,\frac{f_t(\vec{x};\theta)}{g(\vec{x})}g(\vec{x})\,d\vec{x}\right] / \left[\int_{\vec{x}} \mathbf{1}(\vec{x}\in A_t(\theta))\,\frac{f_t(\vec{x};\theta)}{g(\vec{x})}g(\vec{x})\,d\vec{x}\right].$$

This suggests estimating $m_k^t(\theta)$ by

$$\tilde{m}_{k,s}^t(\theta) = \frac{1}{S}\sum_{s=1}^{S} p_{k,s}^t(\theta)$$

where

$$p_{k,s}^t(\theta) = \left[p_k^t(\vec{x}_s;\theta)\,\frac{f_t(\vec{x}_s;\theta)\,\mathbf{1}(\vec{x}_s\in A_t(\theta))}{g(\vec{x}_s)}\right] / \left[\frac{1}{S}\sum_{s=1}^{S}\frac{f_t(\vec{x}_s;\theta)\,\mathbf{1}(\vec{x}_s\in A_t(\theta))}{g(\vec{x}_s)}\right].$$

In addition, to facilitate the optimization procedure, we smooth the indicator function in the above summations in a way similar to the independent probit kernel smoothing in the maximum likelihood procedure. In particular, we estimate $m_k^t(\theta)$ by

$$p_{k,s}^t(\theta) = \left[p_k^t(\vec{x}_s;\theta) \frac{f_t(\vec{x}_s;\theta) \prod_{k=0}^{N_t-2} \prod_{j=1}^{N_t-k-1} \Phi\left(\frac{p_{k,j}^t(\vec{x}_s;\theta) - p_{k,N_t-k}^t(\vec{x}_s;\theta)}{h}\right)}{g(\vec{x}_s)} \right] / \left[\frac{1}{S} \sum_{s=1}^{S} \frac{f_t(\vec{x}_s;\theta) \prod_{k=0}^{N_t-2} \prod_{j=1}^{N_t-k-1} \Phi\left(\frac{p_{k,j}^t(\vec{x}_s;\theta) - p_{k,N_t-k}^t(\vec{x}_s;\theta)}{h}\right)}{g(\vec{x}_s)} \right] \tag{4.3.22}$$

where $p_{k,j}^t$ is the targeted dropout price for bidder j in round k, for auction t, as a function of $\vec{x}_s$ and θ.

In our case, we use the importance density $g(\vec{x};\theta) = f_t(\vec{x};\theta)$, the *untruncated* multivariate normal distribution for auction t. In this way, $f_t(\vec{x}_s;\theta)$ cancels out from both the numerator and denominator in equation 4.3.22.

Asymptotic distribution of $\theta_{\mathbf{SNLS}}$ The asymptotic distribution of θ_{SNLS} is given by:

$$\hat{\Sigma}^{-1/2}\sqrt{T}\left(\theta_{SNLS} - \theta_0\right) \xrightarrow{d} N(0, I)$$

where $\hat{\Sigma} = \hat{\mathcal{A}}^{-1}\hat{\mathcal{B}}\hat{\mathcal{A}}^{-1}$, and

$$\hat{\mathcal{A}} = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=0}^{N_t-2}\left[\frac{\partial \tilde{m}_k^t(\theta)}{\partial\theta}\right]\left[\frac{\partial \tilde{m}_k^t(\theta)}{\partial\theta}\right]',$$

$$\hat{\mathcal{B}} = \frac{1}{T}\sum_{t=1}^{T}\left(\left[\sum_{k=0}^{N_t-2}\frac{\partial \tilde{m}_k^t(\theta)}{\partial\theta}\left(p_k^t - \tilde{m}_k^t(\theta)\right)\right]\left[\sum_{k=0}^{N_t-2}\frac{\partial \tilde{m}_k^t(\theta)}{\partial\theta}\left(p_k^t - \tilde{m}_k^t(\theta)\right)\right]'\right).$$

4.3.4 What are the estimable parameters?

The goal in estimating this model, as in estimating *any* equilibrium auction model, is to recover bidders' valuations from their observed bids (or, in this case, their dropout prices).

We can estimate three sets of parameters:

1. $\theta \equiv (s, t, r_0)$, the standard deviations for the private and common value estimates, and for the common value distribution. To begin with, we will assume that they are the same across all firms. Later on, we can introduce some heterogeneity in both s and t (the latter allowing for differences in the quality of information about the object's value possessed by the different firms). Since these parameters are constrained to be positive, in the results below we report estimates of the *logs* of these quantities.

2. $\bar{a}_i = Z_i'\alpha$– we parameterize the publicly-known mean of bidder i's private valuation of the object to be a linear function of firm-and-object specific covariates.

3. We parameterize m, the mean of the log common value distribution for a given object. Since we don't want to estimate a separate m for each object, we can parameterize m to be a function of object attributes (such as population or size of the spectrum band to be sold).

4.4 Empirical application: the FCC Spectrum Auctions

In the past several years, the FCC has held auctions to allocate Personal Communications Services (PCS) licenses. Each license covers a particular slice of the radio spectrum over a particular geographic area. Licenses were offered both at the MTA and BTA level[16] The data used in this chapter come from the most important spectrum auction, the **MTA broadband PCS auction**, which began on December 5, 1994 and ended on March 13, 1995, after 112 rounds of bidding. 99 MTA licenses were offered– two 30 MHz licenses in most of the 51 MTAs which comprise the US and its territories abroad. (See figure 4.2 for a map showing the numbers, sizes and locations of the MTAs.) In this chapter, we analyze the auctions of 91 of these licenses. 30 firms participated in this auction, and 19 of them eventually won licenses, yielding over $7 billion in government revenue. The so-called A and B block spectra went on sale in the auctions we analyze; figure 4.3 shows the breakdown of the broadband PCS spectrum into the various blocks.

[16]Respectively, *major trading area* and *basic trading area*. This designations are from Rand McNally.

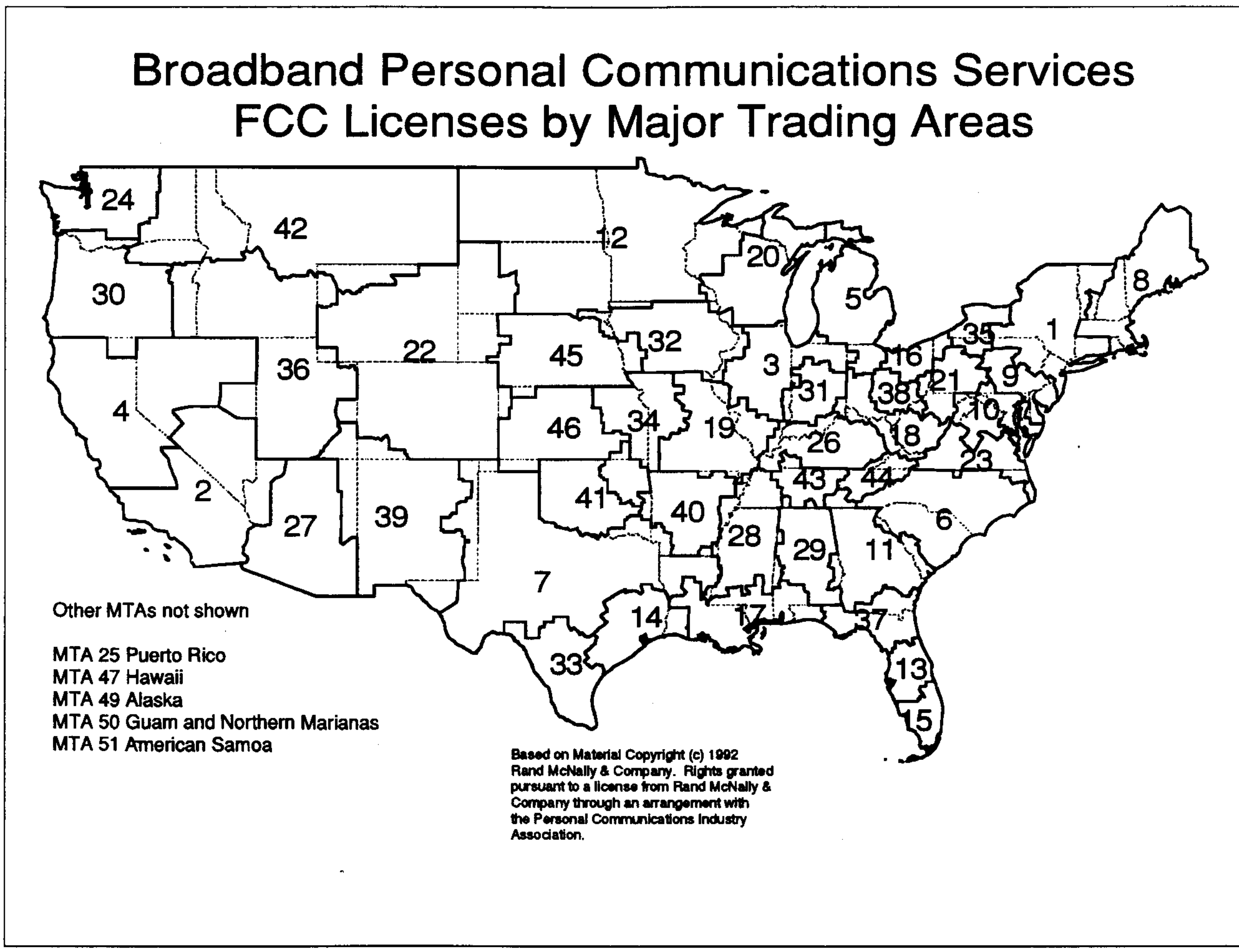


Figure 4.2: MTA Map of the PCS Spectrum auction

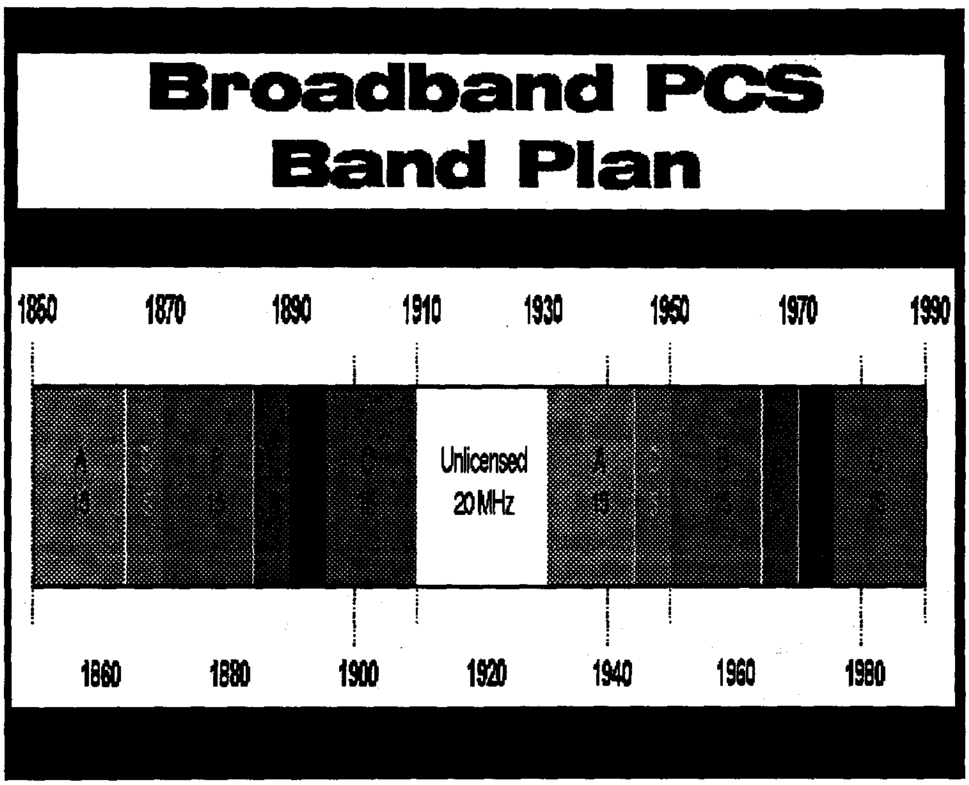


Figure 4.3: Broadband PCS Band Plan

Source: FCC home page

The licenses were allocated using a *simultaneous multiple round auction*, proposed by Paul Milgrom, Robert Wilson, and Preston McAfee. The main features of this auction format is the simultaneity and multiple rounds. Simultaneous auctions of many objects allows bidders to realize cross license synergies, if any exist. The multiple-round format, as explained above, "allows the bidders to react to information revealed in prior rounds, [thus] enabling the bidders to bid more aggressively" (Cramton [27], p. 2). This chapter is concerned about measuring the amount of information revealed in this multiple round auction.

4.4.1 Difficulties with implementing the equilibrium model

Applying the model presented above to the FCC auction data presents some difficulties because the FCC auction format differs from the model in several respects. The most serious of these is that, while the auction is roughly "ascending" in nature, bidders are not required to make irreversible dropout decisions. This violates one assumption of our auction model. Therefore, in order to fit the model to the FCC data, we make some rather strong assumptions about bidders' beliefs concerning the dropout behavior of the other bidders. To begin with, we assign a "dropout price" to bidder j which is the last price at which she was active[17]. In order to estimate the model, we assume that *all remaining bidders also believe that this assigned price is in fact bidder j's dropout price.*

4.4.2 Data and regressors

The data on the auction results from the MTA broadband auction is taken from the FCC's web site (`http://www.fcc.gov`). This data gives us information on the participants and the bids that they submitted during each round on the various licenses. We supplemented this data with market characteristics at the MTA level from the Rand-McNally guide. The cellular presence data came from the *Wireless Market Book* [20] from the Cellular Telephone Industry Association (CTIA). We discuss how we created the dependent variable and the regressors in turn.

[17]We give our definition of "active" below, in the section on the data.

The dropout prices

In section 4.4.1, we explained the difficulties we face in creating dropout prices, since the firms never actually "drop out" from this auction. At that point, we mentioned that we planned to use the last price at which bidder j was active as her ("ex post") dropout price. Now it just remains to define what we mean by an "active" bidder.

The following example will be useful: suppose that there are four bidders (A, B, C, D) and we observe that the last submitted bids for A, B, and C were 10, 20, and 30, respectively. If the price goes up by increments of 5, then, D will win the object at a price of 35 (assuming that his valuation is greater than that).

One simple way would be to assign to each bidder a dropout price equal to his last submitted bid, and assume that the winner's dropout price was greater than or equal to the winning price, i.e., $p_A = 10, p_B = 20, p_C = 30, p_D \geq 35$. This method is inconsistent, because of the gap between the the second-highest dropout price p_C and the lower bound on the highest dropout price p_D. As Milgrom and Weber[57] note, their formulation of the ascending model model reduces to a second-price auction when there are only two bidders left– in this case, these would be C and D. The problem with the above assignment of dropout prices is that the winner – D – doesn't win the object at the "second-price", which is C's dropout price.

To deal with this problem, we assign a dropout price to a given bidder equal to the last submitted bid of the *next* bidder who drops out. In the example above: $p_A = 20, p_B = 30, p_C = 35, p_D \geq 35$. The reason this problem occurs is that the Milgrom and Weber[57] (and Wilson[73]) model assume continuously rising prices and instantaneous dropouts, whereas in the FCC auctions (and probably in most real-life situations) the price ascends by discrete intervals.[18]

[18]To our knowledge, no one has come up with a discrete version of the ascending auction model.

Regressors for common value

The regressors which we use to parameterize MV, the mean of the common value distribution[19], are MTA-level demographic variables which should capture the across-market variation in the value of the licenses. We use (1) population (2) population change, and (3) average household income. Population measures the size of the market, while population change accounts for future market growth. Average income is used as a measure of demand for PCS services. These same variables will also be used to parameterize s, the standard deviation of the common value distribution.

Regressors for private value

These regressors, which will parameterize $\bar{a}_i$, the mean of bidder i's log private value distribution, should capture across-firm variation in the way a license is valued.

The only regressor we use at this point is an indicator of cellular presence in the surrounding area. More precisely, this regressor is a tally of the total number of the BTA's surrounding[20] a particular MTA in which a given firm has cellular presence. For a given firm, these regressors will vary across licenses, and for a given license, these regressors will vary across firms.

Figure 4.4 presents summary statistics for all the variables we use in the analysis.

4.5 Estimation results

Figure 4.5 shows the results for two specifications of the full model, estimated using the simulated nonlinear least squares (SNLS) methodology described above. Results obtained from penalized simulated maximum likelihood for these same specifications are presented in figure 4.6. As mentioned above, we are uncertain about the consistency of the maximum likelihood estimates, but present them for comparison's sake. Furthermore, we were

[19] More precisely, the distribution from which the noisy signals about the object's common value are drawn. The "common value" MV is actually the mean of this distribution.

[20] Firms with substantial cellular coverage in a given market were barred from bidding for PCS spectra in that market.

Figure 4.4: Summary statistics for data variables

Variable	N	mean	StdDev	min	max
Winning prices ($mill)	91[1]	75.87	89.71	4.39	493.5
Population (millions)	91	5.15	4.14	1.15	26.78
Pop'n change (1990-95, %)	91	6.00	3.53	0.40	12.80
Per capita income ('000)	91	15.86	3.71	11.96	20.70
Dropout prices ($mill)	423	53.18	69.28	10^{-5} ??	493.5
Cell. pres	423	0.61	1.28	0	8

[1]: We omitted the observations for: Puerto Rico, Guam, Samoa

unsuccessful in our attempts to estimate the r_0 parameter via MLE.

The **a** and **b** models in both figures4.5 and 4.6 differ in the extent to which log s — the log of the variance on the bidders' common value estimates — is parameterized. From the (simulated) NLS objective function values, the extra covariates clearly improve the performance of the model.

Before discussing the results, we mention a few points about identification. In all the models, the constant term in $\bar{a}$, the mean of the log private value estimates, is not separately identified apart from the constant term in m, the mean of the log common value estimates. Therefore we do not estimate the former.

The goal in estimating this structural model is to recover the parameters of the bidders' underlying valuations of the MTA licenses. These valuations are composed of a common-value component which varies over markets, parameterized by the regressors in m, and a private value component which (in our case) varies over both firms and markets, which is parameterized by the regressors in $\bar{a}$.

A quick comparison of the results in figures 4.5 and 4.6 shows that the parameter estimates change substantially depending on estimation methodology. Furthermore, for the

Figure 4.5: Simulated Nonlinear Least-Squares Estimates
M (number of simulation draws): 100; H (bandwidth)=0.01

Coefficient	Model 1a Estimate	Model 1a Std. Err	Model 1b Estimate	Model 1b Std. Err
Components of log s:				
Constant	2.515	0.354	1.627	0.604
POP (mills)			1.124	0.824
INCOME (per cap.,$'000)			5.852	1.549
Components of log t:				
Constant	3.004	0.315	2.999	0.237
log r_0[a]:				
Constant	0.793	0.318	0.813	0.183
Components of m:				
Constant	10.762	8.306	6.458	14.069
POP (mills)	-0.035	4.301	-0.036	0.081
POP CHANGE (%)	0.212	0.328	0.483	0.511
Components of $\bar{a}$:				
Constant[b]				
CELL PRES	0.009	0.069	0.008	0.019
NLS Obj.Fun.[c]	23002.29		21955.38	
Num of Obs	91		91	

[a] variance of common value distribution
[b] Not separately identified from constant in m
[c] simulated

Figure 4.6: Simulated Maximum Likelihood Estimates
M (number of simulation draws): 100; H (bandwidth)=0.2

Coefficient	Model 2a Estimate	Model 2a Std. Err	Model 2b Estimate	Model 2b Std. Err
Components of log s:				
Constant	0.60875	0.03101	0.00434	0.06085
POP (mills)			-0.04129	0.01968
INCOME (per cap.,$'000)			0.06573	0.00400
Components of log t:				
Constant	-0.69661	0.25432	-0.67536	0.02903
log r_0[a]:				
Constant	0[b]		0	
Components of m:				
Constant	0.99076	0.18696	0.83846	0.27567
POP (mills)	0.09728	0.01977	0.05590	0.04551
POP CHANGE (%)	0.03453	0.01239	0.02596	0.03087
Components of $\bar{a}$:				
Constant[c]				
CELL PRES	-0.01170	0.00881	-0.00017	0.07976
Log Likelihood[d]	138.922		164.767	
Num of Obs	91		91	

[a] variance of common value distribution
[b] fixed
[c] Not separately identified from constant in m
[d] simulated

SNLS results, only the various constant terms are estimated with any reasonable precision; the coefficients attached to the covariates are very imprecisely estimated, and small in magnitude. In particular, in the current specifications, asymmetries across the bidders is captured only by including the CELL PRES covariate in $\bar{a}$, the mean of the distribution from which bidders' private values are drawn. The small and insignificant estimated coefficient on this covariate (0.008 for model 1b) indicates that bidders are largely symmetric given this specifications and our results. This finding has implications on how the equilibrium bid functions for a given bidder changed during the course of the auction. We turn to these bid functions next.

4.5.1 Estimated bid functions

Figure 4.7 shows plots of the estimated (log) bid functions for the winning bidder ("bidder 1", using the indexing in this chapter), in each of the rounds of selected auctions. x_1, the log of his signal, is plotted on the x-axis, while $b_1^k(x_1)$, his log bid functions for rounds $k = 1, \dots, N-1$, are plotted on the y-axis. The units on the y-axis are log($mills).

First note that the log bid functions are linear in the signals; this is due to the log-normality assumption. Secondly, note that the bid functions decrease in slope as the auction progresses, implying that for any given valuation x_1 in the range in which bidder 1 would have won the auction, the targeted dropout price falls as bidders drop out. For example, for auction #89 (Hawaii block B), if $x_1 = 1.5$, then we can read off the graph that bidder 1's targeted log dropout price falls from around $4.1 mills in round 0, to about $2.8 mills in the final round. Furthermore, it is easy to show that a bidder's equilibrium bid functions for any two successive rounds k and $k+1$ intersect at a log price equal to p_k, the log dropout price for round k.[21]

This monotonic change in the slope of the bid functions is characteristic of *symmetric* ascending auctions. As noted above, for our estimates, bidders are essentially symmetric

[21]Essentially, the first $N-k-1$ of the $N-k$ values $(b_1^k)^{-1}(p_k), \dots, (b_{N-k}^k)^{-1}(p_k) = p_k$ which solve the system of $N-k$ equations which define the equilibrium bid functions in round k (see equation 4.2.12) also solve the system of $N-k-1$ equations for round $k-1$, given $p = p_k$.

given our specification. Changes in the slope of the bid function occur because he conditioning events change as the auction progresses, as bidder 1 learns the private signals of the bidders who have dropped out. In a symmetric ascending auction, where no differences exist among his competitors, bidder 1's expected valuation for the object is either increasing or decreasing in each and every of his competitors' private signals. Furthermore, when bidder j remains in the auction, bidder 1 assumes in equilibrium that bidder j's private signal is equal to x_1.[22] Once bidder j drops out, bidder 1 learn x_j and, given symmetry, $x_j < x_1$. Essentially, bidder 1 "plugs" a smaller number x_j into his bid function. This is true for each bidder $j \neq 1$, since bidder 1 wins the auction. This causes the slopes of the successive bid functions to change monotonically as the auction progresses. That the slopes monotonically *decrease* arises if bidder 1's expected valuation for the object is increasing in all his competitors' private signals. This is true for our log-normal model with a common value component, as is clear from equation 4.2.12, in which all the coefficients associated with bidders' private signals (for bidder i, these would be $\frac{s_i^2}{r_i^2 h r_j^2}$ for all bidders $j \neq i$) are positive.

For the asymmetric case, this monotonicity need not hold, even assuming, as in the log-normal model, that bidder 1's expected valuation is increasing in each and every private signal. This is because in every round, bidder 1 not only learns the private signal of the dropout bidder during that round, but also revises his beliefs about the remaining bidders' signals knowing that these bidders are also revising their beliefs upon observing dropout behavior. While the coefficients on all the private signals in bidder 1's expected valuation of the object are still positive, it is not clear whether the new values for the signals "plugged in" during each round are larger or smaller than the old values; therefore, it is unclear how this change in information affects the slope of the bid function. Therefore it is not clear that bidder 1's sequence of targeted dropout prices will change monotonically, as we have found for the symmetric case.

[22] As explained in the very beginning of the paper, this is because, in equilibrium, bidder 1 bids a price $b_1(x_1)$ in which his expected revenue from winning is just equal to $b_1(x_1)$. If he in fact wins at price $b_1(x_1)$, this must mean that bidder j has dropped out at that price, implying that bidder j private signal $x_j = b_j^{-1}(b_1(x_1))$. For the symmetric case, $b_j(\cdot) = b(\cdot)$, $\forall\ j$, so that $x_j = b_j^{-1}(b_1(x_1)) = x_1$.

Nonetheless, there is some tension between our finding of bidders' "pessimism" (as evidenced by the fall in targeted dropout prices) as the auction proceeds with Milgrom and Weber's assertion that information raises *ex ante* expected revenues from an auction. But MW are making a statement about different auction forms (i.e., with common values, English auctions yields higher *ex ante* expected revenue), whereas our pessimism just occurs across the rounds of an English auction: despite it, what MW are saying is that the *ex ante* expected winning price in an English auction exceeds that of a second-price auction, for example.

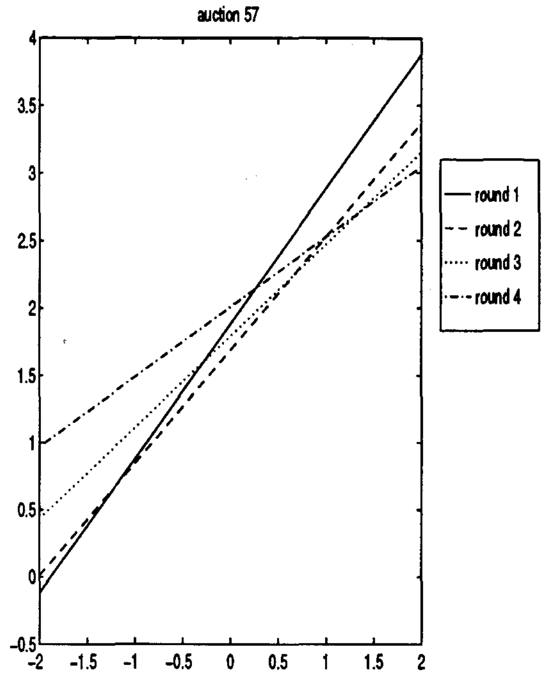
4.6 Conclusions

In this chapter we derive an econometric model of a multi-round ascending (or English) auction. Under the assumption that bidders' private signals are drawn from (not necessarily identical) log-normal distributions, we derive that the vector of log dropout prices observed in a given ascending auction is distributed multivariate truncated normal. We estimate the resulting model using data from the FCC spectrum auctions.

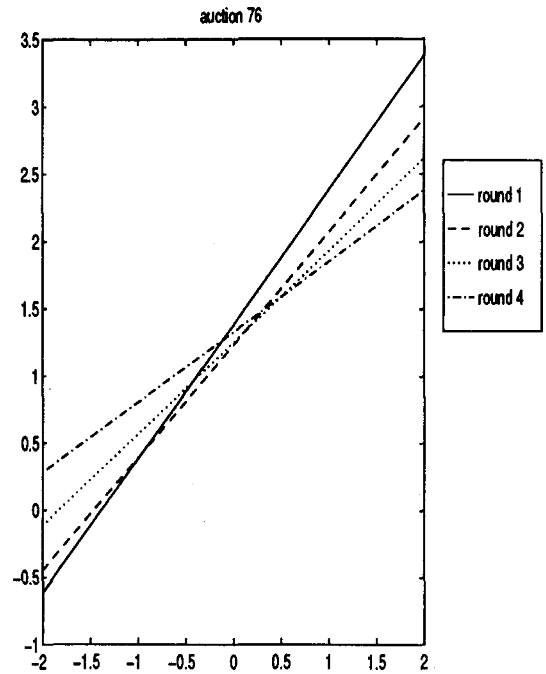
While we would also like to extend our estimating model to incorporate important features of the FCC auctions such as discrete bid increments as well as bidders' uncertainty about their competitors' dropout prices, we are prevented from doing so by theoretical constraints. The problem with making these adjustments is that the resulting models will no longer be equilibrium models– given a dose of uncertainty about dropout prices, for instance, it may well be that "jump bids" and other signaling behavior occur in equilibrium. We are not aware of any theoretical equilibrium auction models both rich enough to accommodate this sort of behavior and simple enough to implement econometrically.

Nevertheless, we believe that the model we have developed here is general enough to be applicable to other ascending auction settings.

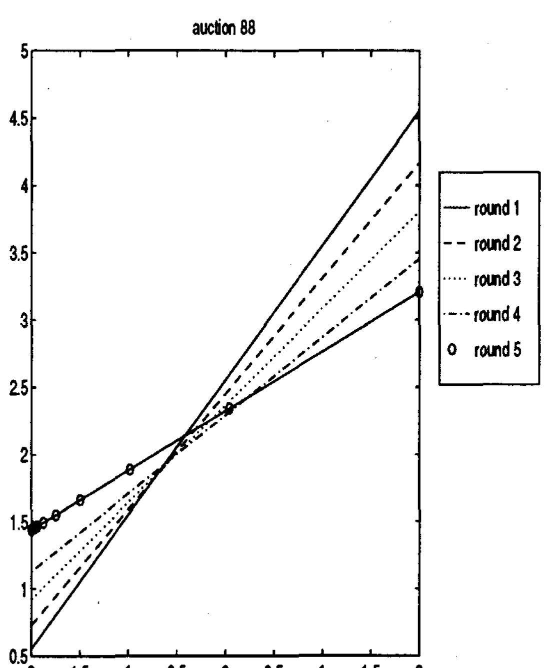
Figure 4.7: Estimated bid functions: a few examples



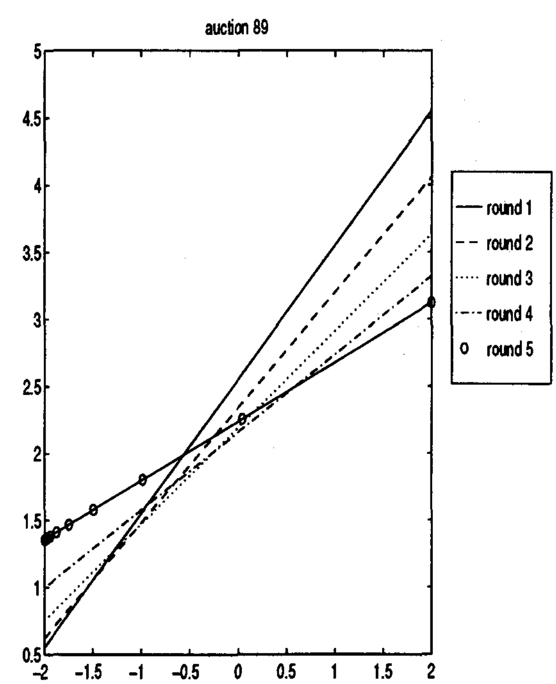
MTA 31, Block B: Indianapolis



MTA 41, Block A: Oklahoma City



MTA 47, Block A: Hawaii



MTA 47, Block B: Hawaii

Appendix A

Appendix to chapter 2

A.1 Calculating long-run aggregate brand purchases and elasticities

Here I describe how I calculated the long-run aggregate brand purchase elasticities reported earlier, in section 2.5. For each household, I —

1. Initialize the purchase history using brands that the household purchased in the 12 weeks prior to 1/1/92.

2. Simulate a 51x1 vector of purchase probabilities $\vec{pp}$ for each brand (as well as the outside good), for week 1, using the same formulas as were used in the estimation (equation 2.3.24). Time-varying covariates (such as price and advertising) are set to their sample averages for each brand.

3. Randomly sample one of the 51 choices, according to the probabilities $\vec{pp}$. Then update the list of brands that the households has experienced. As in the estimated models, I define an experienced brand as one which was chosen in the 12 weeks prior to a given week.

4. Repeat steps 2 and 3 for weeks 2 through 52. This assumes that at most one of the top 50 brands is chosen per week.

5. After 52 weeks, total up this household's purchases for each of the 50 brands, in order to derive the 50x1 vector $\vec{d_h} \equiv (d_{1h}, \ldots, d_{50h})'$.

After running this algorithm for each household, I sum over all households' total (over 52 weeks) purchases for each of the 50 brands, in order to derive the 50x1 vector of aggregate purchases $\vec{D} \equiv (D_1, \ldots, D_{50})' = (\sum_h d_{1h}, \ldots, \sum_h d_{50h})'$. In summing over the households, I weight each household using weights that I derived from the 1992 US *Consumer Expenditure Survey*, in order to control for the unrepresentativeness of my sample households.

The next step is to simulate the *arc* price elasticities of the aggregate purchases of each brand:

$$\epsilon_{ij} \equiv \frac{\%\Delta D_i}{\%\Delta p_j} \approx \frac{(D_i(p_j * (1+\Delta_p); p_{k,k\neq j}) - D_i(p_j; p_{k,k\neq j}))/D_i(p_j; p_{k,k\neq j})}{\Delta_p}$$

where Δ_p is a given percentage change in p_j.

A.2 Comparisons with other recent results

Comparing own-price elasticity estimates At this point it is useful to compare the elasticity estimates recovered from the above models with estimates obtained from recent analyses of this market. How do the long-run elasticities calculated here, which account for households' brand loyalty, compare to Hausman [37] and Nevo's [60] estimates, both of which employed aggregate data and abstracted away from households' brand loyal behavior (so that there was no distinction between short-run and long-run elasticities in the sense used here)? Figure A.1 gives both the short-run and long-run own-price elasticities calculated using the results from models C and E, as well as the estimates reported in the two other papers cited above.

Nevo's average elasticity of -2.95 is somewhat lower that the model E long-run estimate of -3.83. However, Nevo's average is only over 25 brands, whereas the model E average is calculated over 50 brands; the difference in magnitude may be due simply to the fact that the extra brands considered in model E have high own-price elasticities, so that the average elasticities is blown up.

On the individual brand level, however, Nevo's estimate largely agree with my estimates of the short-run elasticities for the experienced households as well as the long-run elasticities, across all the specifications considered in this table. Hausman's estimates, on the other hand, tend to be lower in magnitude than both Nevo's and my estimates. This difference is especially marked for Cheerios: Hausman estimates -1.93, while Nevo and the model E long-run figure largely agree at -3.87 and -3.68, respectively.

The case of Apple-Cinnamon Cheerios Of especial interest are the elasticities for Apple-Cinnamon Cheerios (brand #31), the welfare effects from the introduction of which was the subject of Hausman's [37] recent paper. My estimate of the long-run elasticity (using the model E results) for this brand is -4.08. This is substantially higher in magnitude than Hausman's estimate of -2.17 ([37], table 5.6). The more elastic demand for this brand indicated by my estimate implies that the welfare gain from the introduction of this brand would be lower than Hausman's result would imply.

Alternatively put, the "virtual price" which would set purchases for AC Cheerios to zero would be lower using my elasticity estimate than using Hausman's. Figure A.2 presents calculations of this virtual price using my estimates, as well as Hausman's calculations. Hausman estimates the virtual price of $7.14 under an assumption that prices of all other brands remain constant in the absence of AC Cheerios, with a lower bound of $5.55. My

Figure A.1: Own-price Elasticity Estimates

	Model C results			**Model E results**			**Hausman (1996)**[a]	**Nevo (1997)**[b]
	Short-run [c]		*Long-run*	*Short-run* [d]		*Long-run*[e]		
	Exp. hh's	Inexp. hh's		Exp. hh's	Inexp. hh's			
Average[f]	-2.043	-4.297	-3.16	-2.16	-5.67	-3.83		-2.95
Selected brands:								
KG Corn Flakes	-1.87	-2.15	-2.12	-2.11	-2.63	-2.30	-2.17	-3.39
KG Raisin Bran	-2.17	-3.00	-2.56	-2.37	-3.75	-2.93	-2.07	
KG Rice Krispies	-2.83	-3.70	-3.20	-3.11	-3.43	-3.76	-2.17	-3.25
KG Frosted Flakes	-2.83	-2.89	-2.43	-3.14	-3.43	-2.82		-3.14
GM Cheerios	-3.38	-3.72	-3.32	-3.73	-4.49	-3.87	-1.93	-3.68
GM HN Cheerios	-2.49	-4.30	-3.60	-3.00	-6.40	-3.98	-1.98	
GM AC Cheerios	-2.18	-4.30	-3.18	-2.30	-5.60	-4.08	-2.17	
PT Raisin Bran	-1.86	-3.02	-2.47	-1.98	-3.85	-2.75	-2.62	-2.50

[a]Taken from table 5.6
[b]Taken from tables 8 and 9
[c]Same calculations as reported in top panel of figure 2.14
[d]Same calculations as reported in bottom panel of figure 2.14
[e]Same calculations as reported in figure 2.17.
[f]For Model C and E figures, average over top fifty brands. For Nevo's results, average over 25 brands.

estimates, for all the models considered, tend on average to agree more with Hausman's lower bound estimate: I calculate a virtual price of \$5.85 for the model C results, and \$5.02 using the model E results. My lower estimate of this virtual price would imply lower welfare benefits (measured as the area under the demand curve from zero to observed levels of demand) from the introduction of this brand.

However, the magnitude of this product's welfare contribution depends not only on its own-price elasticity, but also on how well it substitutes with other brands, most notably its "umbrella" parent (Cheerios, brand #2) and sibling (Honey Nut Cheerios, brand #7)[1]. Unfortunately, the cross-price elasticities reported in figure 2.17 are not estimated with enough precision for me to gauge this effect; nevertheless, the small point estimates agree in magnitude with those of Hausman.

Figure A.2: Virtual price calculations, for *Apple Cinnamon Cheerios*
Virtual price p_v is price which sets aggregate long-run demand for Apple Cinnamon Cheerios to be ≤ 5 purchases among the 1010 households during one hypothetical year.

Model	p_0[a]	p_v	$\%\Delta p$[b]
Hausman (1996)[c]:			
Point estimate	\$3.48	\$7.14	+105%
Lower bound	\$3.48	\$5.55	+60%
My estimates:			
Model C	\$3.02	\$5.85 (0.44)[d]	+94%
Model E	\$3.02	\$5.02 (0.37)	+66%

[a] average observed price
[b] $(p_v - p_0)/p_0$
[c] [37], pg. 228
[d] Standard deviation in parentheses. 15 simulations were performed; reported figures are mean and standard deviations over these 15 simulations.

[1] Multigrain Cheerios (brand # 48) was introduced after AC Cheerios.

Appendix B

Appendix to chapter 4

B.1 Details of dropout price density derivation

In this appendix we provide a detailed derivation of the joint density of the dropout prices, given in equation 4.2.16 in the main text.

Let $(V_1, \ldots, V_N, X_1, \ldots, X_N)$ be lognormally distributed. Let $v_i \equiv \ln V_i$ and $x_i \equiv \ln X_i$, $\forall i = 1, \ldots, n$. Assume that $(v_1, \ldots, v_N, x_1, \ldots, x_N)$ are distributed as jointly normal with mean vector $\mathcal{U} = (u_1, \ldots, u_N, u_1^*, \ldots, u_N^*) \equiv (\mu, \mu^*)$ and variance-covariance matrix $\hat{\Sigma} = ((\Sigma, \Sigma_{12}), (\Sigma_{12}', \Sigma^*))'$. Explicit formulas for the elements of the vector $\mathcal{U}$ and the matrix $\hat{\Sigma}$ can be derived from the equations in (4.2.9).

As before, we have

$$E[V_i \mid X_1, \ldots, X_N] = \exp\left(E(v_i \mid x_1, \ldots, x_N) + \frac{1}{2}V(v_i \mid x_1, \ldots, x_N)\right)$$

If we denote the marginal mean-vector and marginal variance-covariance matrix of $(v_i, x_1, \ldots, x_N)$ by

$$\mu_i \equiv (u_i, \mu^*) \quad \text{and} \quad \Sigma_i \equiv \begin{pmatrix} \sigma_i^2 & \sigma_i^{*\prime} \\ \sigma_i^* & \Sigma^* \end{pmatrix} \quad \text{also let} \quad x' \equiv (x_1, \ldots, x_N)$$

Then, using the conditional mean and variance of joint normal random variables (see, for example, Amemiya [3], pg. 3):

$$E(v_i \mid x) = \left(u_i - \mu^{*'}\Sigma^{*-1}\sigma_i^*\right) + x'\Sigma^{*-1}\sigma_i^*$$

and

$$V(v_i \mid x) = \sigma_i^2 - \sigma_i^{*\prime}\Sigma^{*-1}\sigma_i^*.$$

At round k,[1] let $x_d^k \equiv (x_{N-k+1}, \ldots, x_N)$ denote the vector of k valuations for the bidders who have dropped out prior to round k, and $x_r^k \equiv (x_1, \ldots, x_{N-k})$ denote the

[1] $k = 0, \ldots, N-2$. $x_d^0 = \emptyset$.

vector of $(N-k)$ valuations for the bidders who have not yet dropped out as of round k. Analogously, partition Σ^{*-1} into $(\Sigma_{k,1}^{*-1}, \Sigma_{k,2}^{*-1})'$ where $\Sigma_{k,1}^{*-1}$ is a $((N-k)\times n)$ matrix and $\Sigma_{k,2}^{*-1}$ is a $(k \times n)$ matrix. Then the conditional mean function can be re-written as:

$$E(v_i \mid x) = \left(u_i - \mu^* \Sigma^{*-1} \sigma_i^*\right) + x_r^{k'} \Sigma_{k,1}^{*-1} \sigma_i^* + x_d^{k'} \Sigma_{k,2}^{*-1} \sigma_i^*.$$

After substituting the conditional mean and variance formulas into the equations in (4.2.10) and taking the log of both sides, we get the following set of $(N-k)$ linear equations for the $N-k$ bidders active in round k:

$$p \equiv \in P = \left(u_i - \mu^* \Sigma^{*-1} \sigma_i^*\right) + \sigma_i^{*\prime} \Sigma_{k,2}^{*-1\prime} x_d^k + \sigma_i^{*\prime} \Sigma_{k,1}^{*-1\prime} x_r^k + \frac{1}{2}\left(\sigma_i^2 - \sigma_i^{*\prime} \Sigma^{*-1} \sigma_i^*\right) \quad \text{(B.1.1)}$$

for $i = 1, 2, \cdots, N-k$. Equations (4.2.12) in the main text are derived from (B.1.1) by substituting in the explicit formulas for $(u_i, \mu^*, \Sigma^{*-1}, \sigma_i^*, \Sigma_{k,1}^{*-1}, \Sigma_{k,2}^{*-1})$ which we can derive from the equations in (4.2.9).

Deriving $P_k = f(x_{N-k}; x_{N-k+1} \dots x_N)$, the "regression equation" for round k dropout price p_k If we let l_k be the $(N-k)\times 1$ vector of 1's, $\mu_k = (u_1, \dots, u_{N-k})'$, $\Gamma_k = (\sigma_1^2, \dots, \sigma_{N-k}^2)'$, $\Lambda_k = (\sigma_1^*, \dots, \sigma_{N-k}^*)'$, then we could rewrite the above system of linear equations (B.1.1) as

$$p \times l_k = \frac{1}{2}\left(\Gamma_k - diag(\Lambda_k \Sigma^{*-1} \Lambda_k') + \Lambda_k \Sigma_{k,2}^{*-1\prime} x_d^k\right) + \mu_k - \Lambda_k \Sigma^{*-1} \mu^* + \Lambda_k \Sigma_{k,1}^{*-1\prime} x_r^k. \quad \text{(B.1.2)}$$

Solving out for the x_r^k, we obtain the set of $(N-k)$ log-inverse bid functions at stage k:

$$x_r^k = \left[\left(\Lambda_k \Sigma_{k,1}^{*-1\prime}\right)^{-1} l_k\right] \times p - \frac{1}{2}\left(\Lambda_k \Sigma_{k,1}^{*-1\prime}\right)^{-1} \quad \text{(B.1.3)}$$

$$\left(\Gamma_k - diag(\Lambda_k \Sigma^{*-1} \Lambda_k') + \Lambda_k \Sigma_{k,2}^{*-1\prime} x_d^k + 2\mu_k - 2\Lambda_k \Sigma_k^{*-1} \mu^*\right). \quad \text{(B.1.4)}$$

Next let us introduce some shorthand notation:

$$\mathcal{A}^k \equiv \left(\Lambda_k \Sigma_{k,1}^{*-1\prime}\right)^{-1} l_k$$

and

$$\mathcal{B}^k \equiv \frac{1}{2}\left(\Lambda_k \Sigma_{k,1}^{*-1\prime}\right)^{-1}\left(\Gamma_k - diag(\Lambda_k \Sigma^{*-1} \Lambda_k') + \Lambda_k \Sigma_{k,2}^{*-1\prime} x_d^k + 2\mu_k - 2\Lambda_k \Sigma^{*-1} \mu^*\right),$$

both of them $(N-k)$-dimensional vectors. Then the inverse bid functions in (B.1.3) can be rewritten:

$$x_r^k = \mathcal{A}^k \times p - \mathcal{B}^k,$$

or, each equation singly:

$$x_{ri}^k = \mathcal{A}_i^k \times p - \mathcal{B}_i^k, \quad \text{for } i = 1, \dots, N-k, \quad \text{(B.1.5)}$$

where $\mathcal{A}_i^k$ and $\mathcal{B}_i^k$ are the ith elements of the vectors $\mathcal{A}^k$ and $\mathcal{B}^k$.[2]

As described in the main text, in round k we keep the $(N-k)$-th equation from the system in (B.1.5); this is the equation $x_r^k{}_{N-k}$ in the notation of (B.1.5). This equation gives us x_{N-k} as a function of p and $x_{N-k+1}, \dots, x_N$. As described before, we do this for rounds $k = 0, \dots, N-2$, and end up with $N-1$ analogous equations for $x_2, \dots, x_N$: we can denote these by $x_r^0{}_N, \dots, x_r^k{}_{N-k}, \dots, x_r^{N-2}{}_2$.

From the data, we observe the vector of dropout prices $\mathcal{P} \equiv (p_0, p_1, \dots, p_{N-2})$. We can invert each of these equations described in the previous paragraph singly in order to derive an equation for p_k, the log dropout price for round k, as a linear function of $x_{N-k}, \dots, x_N$, for $k = 0, \dots, N-2$. We will get

$$p_k = \frac{x_{N-k} + \mathcal{B}_{N-k}^k}{\mathcal{A}_{N-k}^k}, \quad \text{for } k = 0, \dots, N-2. \tag{B.1.6}$$

Since the $\mathcal{P}$ vector is a linear transformation of the x vector, it is also multivariate normal.

For purposes of deriving the joint density, it is useful to look at these equations describing the dropout prices in some more detail. To simplify notation, label bidders with $N, N-1, \dots, 1$ in the order they drop out. Further decompose $\mathcal{B}^k$ into

$$\mathcal{C}^k \equiv \frac{1}{2}\left(\Lambda_k \Sigma_{k,1}^{*-1'}\right)^{-1}\left(\Gamma_k - diag(\Lambda_k \Sigma^{*-1}\Lambda_k') + 2\mu_k - 2\Lambda_k \Sigma^{*-1}\mu^*\right)$$

and

$$\mathcal{D}^k \equiv \frac{1}{2}\left(\Lambda_k \Sigma_{k,1}^{*-1'}\right)^{-1}\left(\Lambda_k \Sigma_{k,2}^{*-1'}\right)$$

Then the equations for $\mathcal{P}$ in terms of the x's in (B.1.6) can be rewritten as

$$\begin{aligned} p^0 &= \frac{\mathcal{C}_N^0}{\mathcal{A}_N^0} + \frac{x_N}{\mathcal{A}_N^0} \\ p^i &= \frac{\mathcal{C}_{N-i}^i}{\mathcal{A}_{N-i}^i} + \frac{x_{N-i}}{\mathcal{A}_{N-i}^i} + \frac{\mathcal{D}_{N-i}^i(x_{N-i+1}, \dots, x_N)'}{\mathcal{A}_{N-i}^i} \quad i = 1, \dots, N-2 \end{aligned} \tag{B.1.7}$$

Again, introduce more shorthand notation:

$$\mathcal{F} = \left(\frac{\mathcal{C}_N^0}{\mathcal{A}_N^0}, \dots, \frac{\mathcal{C}_{N-2}^2}{\mathcal{A}_{N-2}^2}\right)', \quad \mathcal{G}_i = \left(\underbrace{0, \dots, 0}_{N-i-1}, \frac{1}{\mathcal{A}_{N-i}^i}, \frac{\mathcal{D}_{N-i}^i}{\mathcal{A}_{N-i}^i}\right), \ i = 0, 1, \dots, N-2$$

Let $\mathcal{G} = (\mathcal{G}_0', \dots, \mathcal{G}_{N-2}')'$

Given this notation, the system of equations describing the dropout prices (B.1.7) can be very succinctly written as:

$$\mathcal{P} = \mathcal{F} + \mathcal{G}(x_2, \dots, x_N)'.$$

[2]The original bid function in its level can then be written as $b_i(X_i) = \exp\left(\frac{x_i + \mathcal{B}_i^k}{\mathcal{A}_i^k}\right) = \exp\left(\frac{\mathcal{B}_i^k}{\mathcal{A}_i^k}\right) X_i^{\frac{1}{\mathcal{A}_i^k}}$.

We denote the model parameters by θ. Suppose we also have a set of covariates z. The above mean and variance of observed dropout prices, together with the coefficients and constant terms in the truncation conditions, are all functions of the parameter vector θ and the covariates z.

Mean of the bidder evaluation x	$\mu^* = \mu^*(\theta; z)$
Var-Cov of the bidder evaluation x	$\Sigma^* = \Sigma^*(\theta; z)$
Jacobian of transforming x into $\mathcal{P}$	$\mathcal{G} = \mathcal{G}(\theta; z)$
Constant terms in the transformation of x into $\mathcal{P}$	$\mathcal{F} = \mathcal{F}(\theta; z)$

The mean vector μ_p and variance-covariance matrix Σ_p of the vector of dropout prices are:

$$\mu_p(\theta, z) = \mathcal{F}(\theta, z) - \mathcal{G}(\theta, z)\,\mu^*(\theta, z)$$
$$\Sigma_p(\theta, z) = \mathcal{G}(\theta, z)\,\Sigma^*(\theta, z)\,\mathcal{G}(\theta, z)'$$

and therefore the joint density of these prices is

$$f(\mathcal{P}, \theta, z) \equiv (2\pi)^{-\frac{n}{2}} |\Sigma_p(\theta, z)|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathcal{P} - \mu_p(\theta, z))'\Sigma_p(\theta, z)^{-1}(\mathcal{P} - \mu_p(\theta, z)).\right]$$

Equilibrium truncation conditions The equilibrium truncation conditions can also be explicitly expressed in terms of the shorthand notation introduced above.

At each stage k of the auction, the targeted dropout prices of the remaining bidders for that round must be higher than the dropout price at that round, in other words, $p^k \equiv p^k_{N-k} < p^k_i$, for $i = 1, \ldots, N-k-1$, $k = 0, \ldots, N-2$, where p^k_i denotes the targeted dropout price of bidder i in stage k. Note that $x = \mathcal{G}^{-1}(\mathcal{P} - \mathcal{F})$. Let e^N_i be the ith column of a $n \times n$ identity matrix, and let $\mathcal{E}^N_k = (e^N_{N-k+1}, \ldots, e^N_N)'$, then $x_i = e^{N\prime}_i \mathcal{G}^{-1}(\mathcal{P} - \mathcal{F})$, and $(x_{N-k+1}, \ldots, x_N)' = \mathcal{E}^N_k \mathcal{G}^{-1}(\mathcal{P} - \mathcal{F})$. Since $p^k_i = \frac{\mathcal{C}^i_k}{\mathcal{A}^i_k} + \frac{x_i}{\mathcal{A}^k_i} + \frac{\mathcal{D}^k_i(x_{N-k+1}, \ldots, x_N)}{\mathcal{A}^k_i}$, then $p^k_i = \frac{\mathcal{C}^k_i}{\mathcal{A}^k_i} + \frac{e^{N\prime}_i \mathcal{G}^{-1}(\mathcal{P} - \mathcal{F})}{\mathcal{A}^k_i} + \frac{\mathcal{D}^k_i \mathcal{E}^k_i \mathcal{G}^{-1}(\mathcal{P} - \mathcal{F})}{\mathcal{A}^k_i}$, the constraint that $p^k \equiv p^k_{N-k} < p^k_i$ can thus be rewritten as

$$\begin{aligned}
&\left(e^{N\;\prime}_{N-k} - \frac{e^{N\prime}_i \mathcal{G}^{-1}}{\mathcal{A}^k_i} - \frac{\mathcal{D}^k_i \mathcal{E}^k_i \mathcal{G}^{-1}}{\mathcal{A}^k_i}\right)\mathcal{P} < \frac{\mathcal{C}^k_i}{\mathcal{A}^k_i} - \frac{e^{N\prime}_i \mathcal{G}^{-1}\mathcal{F}}{\mathcal{A}^k_i} - \frac{\mathcal{D}^k_i \mathcal{E}^k_i \mathcal{G}^{-1}\mathcal{F}}{\mathcal{A}^k_i}, \\
&\text{For} \quad \left(\begin{array}{c} i = 1, \ldots, N-k-1 \\ k = 0, \ldots, N-2 \end{array}\right)
\end{aligned} \qquad \text{(B.1.8)}$$

These truncation conditions, expressed in equation (B.1.8), involve model parameters. Both the coefficients and the constant terms in equation (B.1.8) are functions of parameters and covariates, i.e.,

$$\mathcal{A}^k_i = \mathcal{A}^k_i(\theta, z) \quad \mathcal{D}^k_i = \mathcal{D}^k_i(\theta, z) \quad \mathcal{C}^k_i = \mathcal{C}^k_i(\theta, z).$$

This violates regularity conditions for maximum likelihood estimation.

Joint density Given these truncation conditions, the joint density of the observed vector of dropout prices $\mathcal{P}$, given the parameter vector θ and the covariates z, is given by

$$L(\mathcal{P} \mid \theta, z) = \begin{cases} \dfrac{f(\mathcal{P}, \theta, z)}{\int_{\mathcal{A}} f(\mathcal{P}, \theta, z)\, d\mathcal{P}} & \text{if truncation conditions hold.} \\ 0 & \text{otherwise.} \end{cases}$$

where $\mathcal{A}$ denotes the event indicated by conditions (B.1.8) when $\mathcal{P}$ is treated as a random vector generated from the multivariate normal density $f(\mathcal{P}, \theta)$.

Bibliography

[1] D. Ackerberg. Advertising, learning, and consumer choice in experience good markets: An empirical investigation. mimeograph, Boston University, January 1996.

[2] D. Ackerberg. Empirically distinguishing informative and prestige effects of advertising. mimeograph, Boston University, January 1996.

[3] Takeshi Amemiya. *Advanced Econometrics.* Harvard University Press, 1985.

[4] S. Anderson, A. DePalma, and J. Thisse. *Discrete Choice Theory of Product Differentiation.* MIT Press, 1992.

[5] R. Andrews and T.C. Srinivasan. Studying consideration effects in empirical choice models using scanner panel data. *Journal of Marketing Research*, 32:30–41, February 1995.

[6] *1996 Annual Report.* Astra-Merck Inc., 1996.

[7] Kyle Bagwell. Informational Product Differentiation as a Barrier to Entry. *International Journal of Industrial Organizaation*, 8(2):207–223, June 1990.

[8] J. Bain. *Barriers to new competition.* Harvard University Press, 1956.

[9] F. Bass, V. Mahajan, and E. Muller. New product diffusion models in marketing. *Journal of Marketing*, 54:1–26, 1990.

[10] M. Ben-Akiva and Brian Boccara. Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, pages 9–24, 1995.

[11] Moshe Ben-Akiva and S. Lerman. *Discrete Choice Analysis.* MIT Press, 1985.

[12] E. Berndt, L. Bui, D. Reiley, and G. Urban. The roles of marketing, product quality and price competition in the growth and composition of the us anti-ulcer drug industry. Working paper, 1994.

[13] S. Berry. Estimating discrete choice models of product differentiation. *Rand Journal of Economics*, 25(2):242–262, 1994.

[14] S. Berry, M. Carnall, and P. Spiller. Airline hubs: Costs, markups, and the implications of customer heterogeneity. NBER Working paper #5561, May 1996.

[15] S. Berry, J. Levinsohn, and A. Pakes. Differentiated Products Demand Systems from a Combination of Micro and Macro Data: Autos Again. Yale University, mimeo., 1997.

[16] S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, 63:841–890, July 1995.

[17] R. S. Bond and D.F. Lean. Sales Promotion and Product Differentiation in Two Prescription Drug Markets. Technical report, Federal Trade Commission, February 1977.

[18] T. Bresnahan, S. Stern, and M. Trajtenberg. Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *Rand Journal of Economics*, pages S17–S44, 1997.

[19] N. Scott Cardell. Variance components structures for the extreme value and logistic distributions with applications to models of heterogeneity. *Econometric Theory*, pages 185–213, 1997.

[20] *Wireless Market Book.* Cellular Telephone Industry of America, 1996.

[21] Cereals to pare ad plans. *Advertising Age*, June 1, 1996, pg. 1.

[22] G. Chamberlain. Panel data. In Z. Griliches and M. Intriligator, editors, *Handbook of Econometrics, Vol. 2.* North Holland, 1984.

[23] J. Chiang, S. Chib, and C. Narasimhan. Markov chain monte carlo and models of consideration set and parameter heterogeneity. John Olin School of Business Working Paper 96-07, Washington University, St. Louis.

[24] P. Chintagunta, E. Kyriazidou, and J. Perktold. Panel data analysis of household brand choice. University of Chicago, mimeo, November 1997.

[25] Andrea Coscelli. Entry of new drugs and doctors' prescriptions. Stanford University, GSB, June 1997.

[26] Andrea Coscelli. Market shares in drug markets: Doctors' or patients' preferences? Stanford University, GSB, June 1997.

[27] P. Cramton. The PCS spectrum auctions: An early assessment. Forthcoming in *Journal of Economics and Management Strategy*, 1997.

[28] Morris DeGroot. *Optimal Statistical Decisions.* McGraw-Hill Book Company, 1970.

[29] J. Deighton, C. Henderson, and S. Neslin. The effects of advertising on brand switching and repeat purchasing. *Journal of Marketing Research*, pages 28–43, 1994.

[30] J. Dubin and D. McFadden. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, 52(2):345–362, 1984.

[31] T. Erdem and M. Keane. Decision-making under uncertainty : Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science*, pages 1–20, 1996.

[32] J. Ferguson. *Advertising and Competition: Theory, Measurement, and Fact.* Ballinger Publishing Company, 1974.

[33] P. Goldberg. Product differentiation and oligopoly in international markets: The case of the US automobile industry. *Econometrica*, 63:891–951, July 1995.

[34] D.Y. Graham. Treatment of Peptic Ulcers caused by Helicobacter Pylori [edi torial; comment]. *New England Journal of Medicine*, 328(5):349–50, Feb 1993.

[35] G. Grossman and C. Shapiro. Informative advertising with differentiated products. *Review of Economic Studies*, 51:63–81, 1984.

[36] P. Guadagni and J. Little. A logit model of brand choice calibrated on scanner data. *Marketing Science*, 2(3):203–238, 1983.

[37] J. Hausman. Valuation of new goods under perfect and imperfect competition. In T. Bresnahan and R. Gordon, editors, *The Economics of New Goods.* University of Chicago Press, 1996.

[38] J. Heckman. Statistical models for discrete panel data. In C. Manski and D. McFadden, editors, *Structural Analysis of Discrete Data.* MIT Press, 1981.

[39] H. Hong. Characterization of equilibria in asymmetric open auctions. mimeo., 1997.

[40] H. Hong. Non-regular meximum likelihood estimation in auction, job search and production frontier models. mimeo., 1997.

[41] *Marketing Fact Book.* Information Resources, Inc. years 1983, 1988, 1992.

[42] V. Kanetkar, C. Weinberg, and D. Weiss. Price sensitivity and television advertising: Some empirical findings. *Marketing Science*, pages 359–371, 1992.

[43] A. Kaul and D. Wittink. Empirical generalizations about the impact of advertising on price sensitivity and price. *Marketing Science*, pages G151–G160, 1995.

[44] J. Kennan. Simultaneous equations bias in disaggregated econometric models. *Review of Economic Studies*, 56:151–156, 1989.

[45] P. Klemperer. Auctions with almost common values: The "wallet game" and its applications. mimeo.

[46] J. J. Laffont, H. Ossard, and Q. Vuong. Econometrics of first-price auctions. *Econometrica*, 1995.

[47] R. Lal and C. Narasimhan. The inverse relationship between manufacturer and retailer margins: A theory. *Marketing Science*, 15(2):132–151, 1996.

[48] James Lattin and J. H. Roberts. Modeling the Role of Risk-Adjusted Utility in the Diffusion of Innovations. Stanford University Graduate School of Business Working Paper 1019, 1989.

[49] *Ad$ Summary.* Leading National Advertisers. years 1990-1993.

[50] G. Mankiw and M. Whinston. Free entry and social inefficiency. *Rand Journal of Economics*, 1985.

[51] D. McFadden. Modelling the choice of residential location. In A. Karlquist et. al., editor, *Spatial Interaction Theory and Residential Location.* North Holland Pub. Co., 1978.

[52] D. McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5):995–1026, 1989.

[53] D. McFadden and P. Ruud. Estimation by simulation. *Review of Economics and Statistics*, pages 591–607, 1995.

[54] Daniel L. McFadden. Lectures on simulation-assisted statistical inference. papers presented at the EC-Squared Conference, Florence, Italy, December 1996.

[55] *TV Dimensions.* Media Dynamics, Inc. Years 1993, 1995, 1997.

[56] Medical Economics Co., editor. *Physicians' Desk Reference (Print ed.).* Oradell, N.J., 1997.

[57] P. Milgrom and R. Weber. A theory of auctions and competitive bidding. *Econometrica*, 50:1089–1122, 1982.

[58] T. Mroz and D. Guilkey. Discrete factor approximations for use in simultaneous equation models with both continuous and discrete endogenous variables. mimeo, University of North Carolina, 1991.

[59] P. Nelson. Advertising as information. *Journal of Political Economy*, pages 729–755, 1974.

[60] A. Nevo. Measuring market power in the RTE cereal industry. University of California, Berkeley, mimeo, 1997.

[61] H. Paarsch. Deciding between the common and private value paradigms in empirical models of auctions. *Journal of Econometrics*, 51:191–215, 1992.

[62] Martina Pertile. ZANTAC. Il Successo ... e' "un Qualcosa di Piu"'. M.Sc. Thesis, CUOA—Consorzio Universitario per gli studi di Organi zzazione Aziendale, 1995.

[63] Charles Phelps. Diffusion of information in medical care. *Journal of Economic Perspectives*, 1992.

[64] M. Roberts and L. Samuelson. An empirical analysis of dynamic, nonprice competition in an oligopolistic industry. *Rand Journal of Economics*, pages 200–220, 1988.

[65] Richard Schmalensee. Product Differentiation Advantages of Pioneering Brands. *American Economic Review*, 72(3):349–365, June 1982.

[66] C. Shapiro. Advertising as a barrier to entry? Federal Trade Commission Bureau of Economics Working Paper #74, 1982.

[67] R. Steiner. A dual-stage approach to the effects of brand advertising on competition and price. In John Cady, editor, *Marketing and the Public Interest.* Marketing Science Institute, 1977.

[68] Scott Stern. The demand for pharmaceuticals. Stanford University mimeo., 1995.

[69] L. Sterne. *The Life and Opinions of Tristram Shandy.* Oxford Paperbacks. first published 1759–67.

[70] G. Stigler. The economics of information. *Journal of Political Economy*, 1964.

[71] John Sutton. *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration.* MIT Press, 1991.

[72] J. Swait. Probabilistic choice set generation in transportation demand models. Civil Engineering doctoral dissertation, MIT, 1984; especially chaps. 2,3.

[73] R. Wilson. Sequential equilibria of asymmetric ascending auctions. class notes, 1995.

[74] F. Wolak. The welfare impacts of competitive telecommunications supply. *Brookings Paper on Economic Activity: Microeconomics*, pages 269–340, 1996.

...
y quedaba [...]
sólo un rumor
cada vez más distante
hasta que todo lo que pudo ser
se convirtió en silencio.
— Pablo Neruda